Machine Learning

CSE 6363 (Fall 2016)

Lecture 1 Introduction

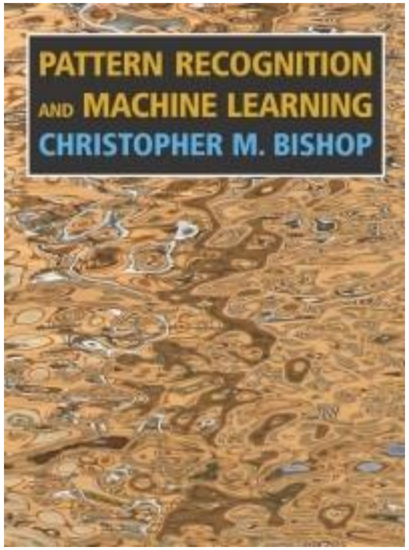
Heng Huang, Ph.D.
Department of Computer Science and Engineering

Administration

- Lecture
  - When: Tue & Thu 2pm ~ 3:20pm
  - Where: WH 221
  - Lecturer: Heng Huang (Office ERB 533) heng@uta.edu
  - Office hour: Tue & Thu 3:20pm ~ 5:00pm (Anytime is ok, if I am in office)
  - Home page: **http://ranger.uta.edu/~heng/CSE6363.html**

Study Materials

- Require Experiences for Course:
  - Mathematics (calculus, algebra, statistics)
  - Algorithms
- Textbook:
  - *Pattern Recognition and Machine Learning, Christopher M. Bishop, 2006.*

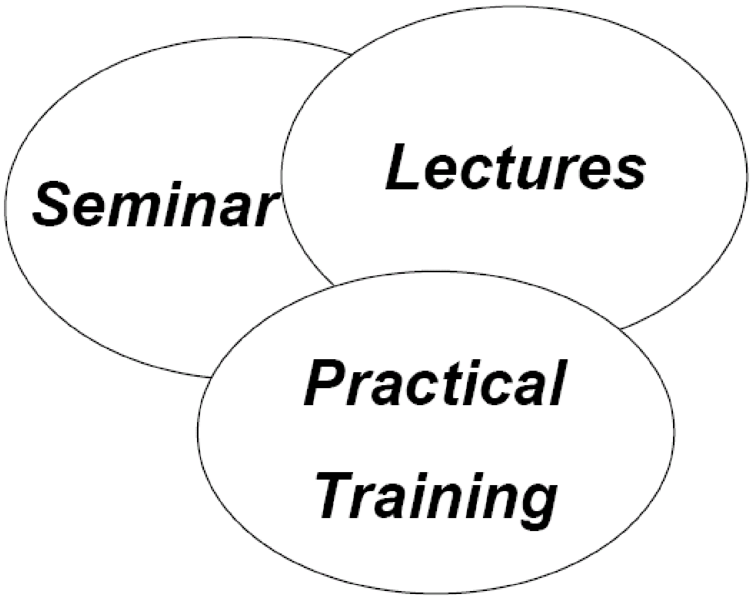


Study Materials

- Some good books for reading:
  - *Machine Learning*, Tom Mitchell.
  - *Pattern Classification 2nd edition* by Richard Duda, Peter Hart, & David Stork.
  - *Information Theory, Inference, and Learning Algorithms*, David Mackay.
  - *Elements of statistical learning*, Friedman, Hastie, & Tibshirani. Springer, 2009.

Course Organization

- Lectures
  - Methodology of machine learning
  - Math behind the machine learning
  - Application: computer vision, bioinformatics, data mining, finance, biomedical image computing, etc.
- Practical Training
  - Homework
  - Final project
- Seminar
  - Read papers
  - Oral presentation



Course Organization

- Grading
  - Three/Four homework sets: 45%
  - Class presentation: 25%
  - Final project: 30%
- No exams
- Homework
  - The homework will be notified in class and at website
  - Two to three weeks for each homework set
  - The homework sets are programming projects
  - Start early, Start early, Start early, Start early, Start early

Course Organization

- Presentation:
  - Reading at least one paper (help you study the applications of machine learning) which will be assigned at the end of September
  - Leading discussion in class (20 mins)
- Project:
  - You select one topic from a project list
  - What do you like? Please give me your background
  - Show your project results in Dec. (5-10 mins)

What's Learning?

- "Learning denotes changes in a system that ... enable a system to do the same task more efficiently the next time." --Herbert Simon
- "Learning is constructing or modifying representations of what is being experienced." --Ryszard Michalski
- "Learning is making useful changes in our minds." --Marvin Minsky
- Summary:
  - Using past experiences to improve future performance
  - For a machine, experiences come in the form of data
  - What does it mean to improve performance?
    - Learning is guided by an objective, associated with a particular notion of loss to be minimized (or, equivalently, gain to be maximized).

Machine Learning vs Pattern Recognition

- ML has origins in Computer Science
- PR has origins in Engineering
- They are different facets of the same field
- So far ML society is more successful
- Most likely ML will cover PR

- Other major related research areas: computer vision, bioinformatics, data mining, information retrieval

Very Brief History

- Studied ever since computers were invented (e.g. Samuel's checkers player)
- Coined as "machine learning" in late 70s - early 80s
- Very active research field, several yearly conferences (e.g., ICML, NIPS), major journals (e.g. Journal of Machine Learning Research, Machine Learning)
- Other related conferences (CVPR, ICCV, AAAI, IJCAI, ECML, ECCV, KDD, UAI, COLT), related important journal (PAMI, TKDE)
- The time is right to start studying in the field!
  - Recent progress in algorithms and theory
  - Growing flood of on-line data to be analyzed
  - Computational power is available
  - Growing demand for industrial applications

Why Do Machine Learning?

- Why machine learning?
  – We need computers to make informed decisions on new, unseen data.
  – Often it is too difficult to design a set of rules “by hand”.
  – Machine learning is about automatically extracting relevant information from data and applying it to analyze new data.

- Let’s study some real research applications using machine learning

Visual Object Categorization

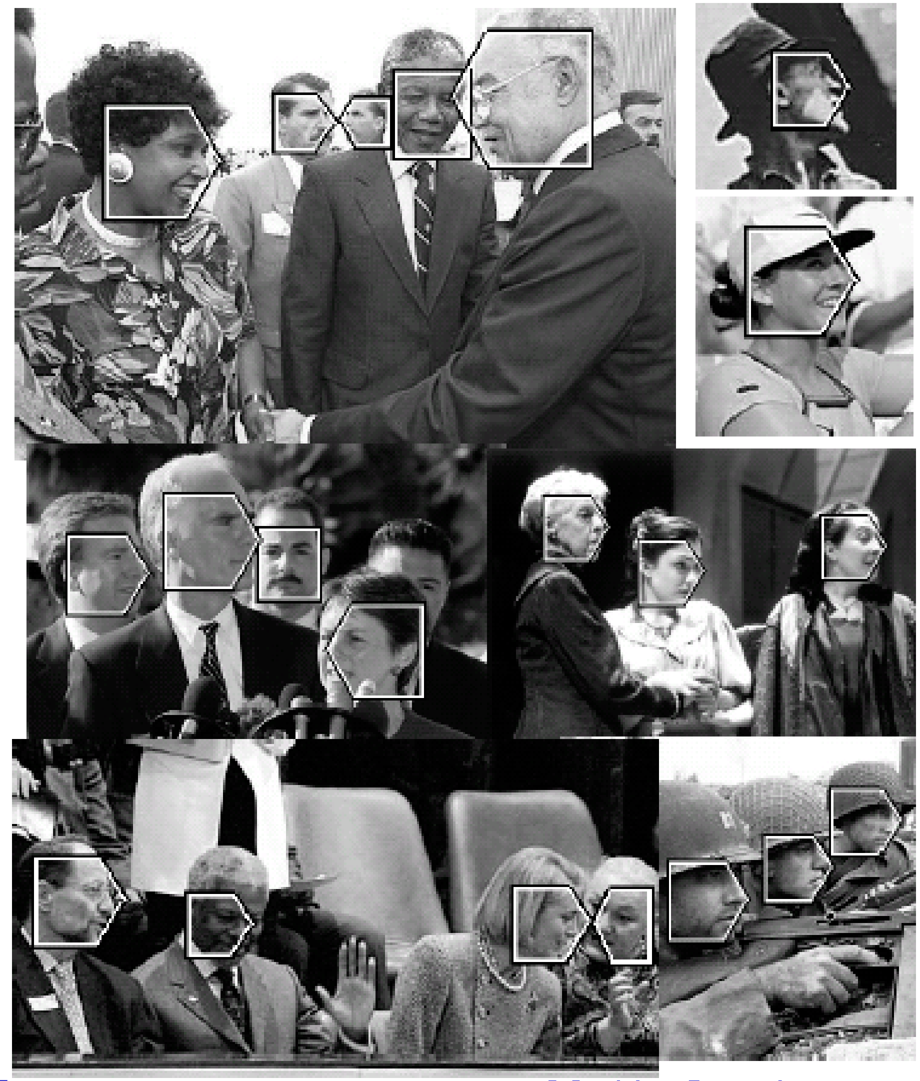
We are given categories for these images:

From ETH database of object categories, [Leibe & Schiele 2003]

What are these?

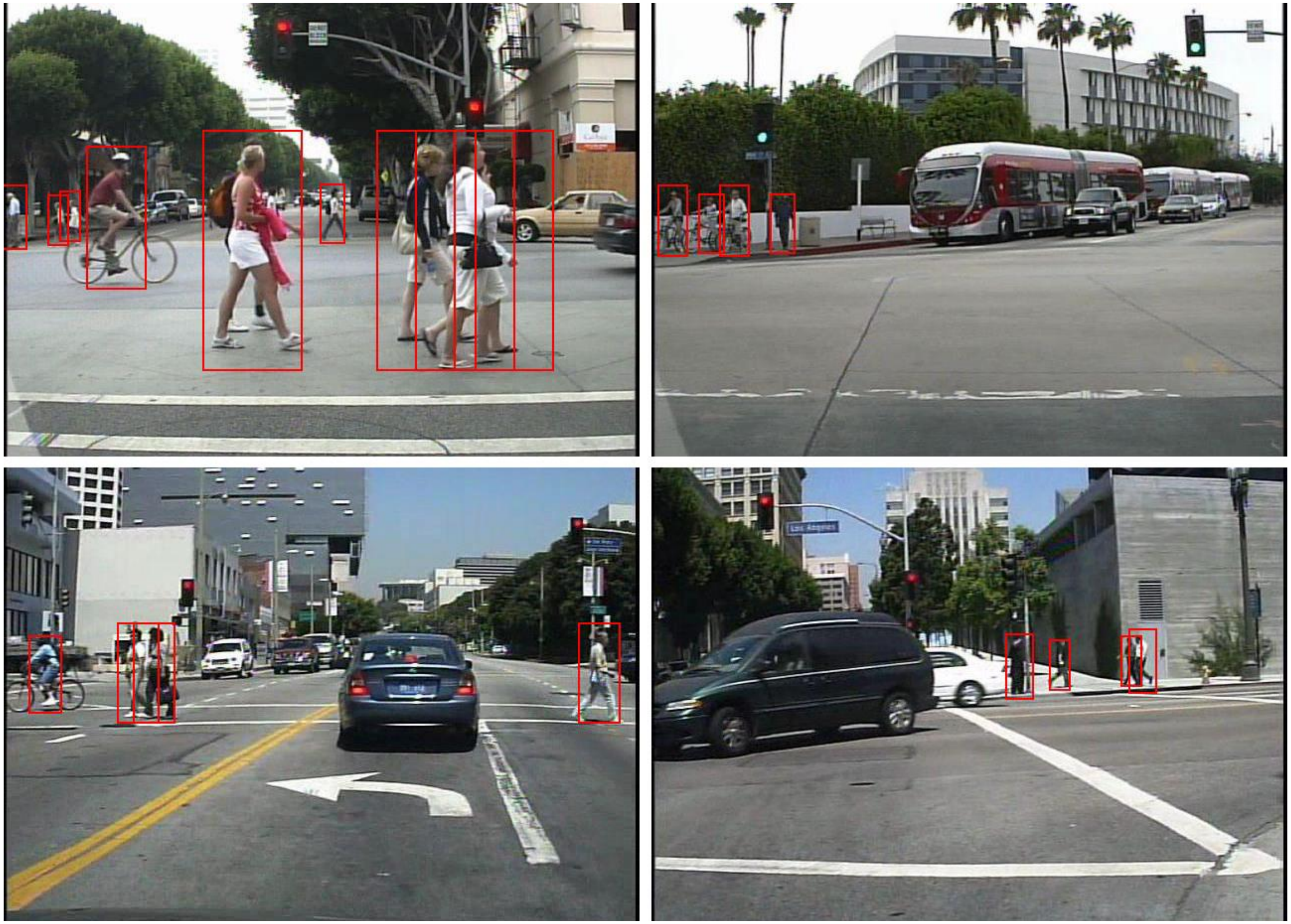
- A classification problem: predict category y based on image x.
- Little chance to "hand-craft" a solution, without learning.
- Applications: robotics, HCI, web search (a real image Google...)

Object Detection



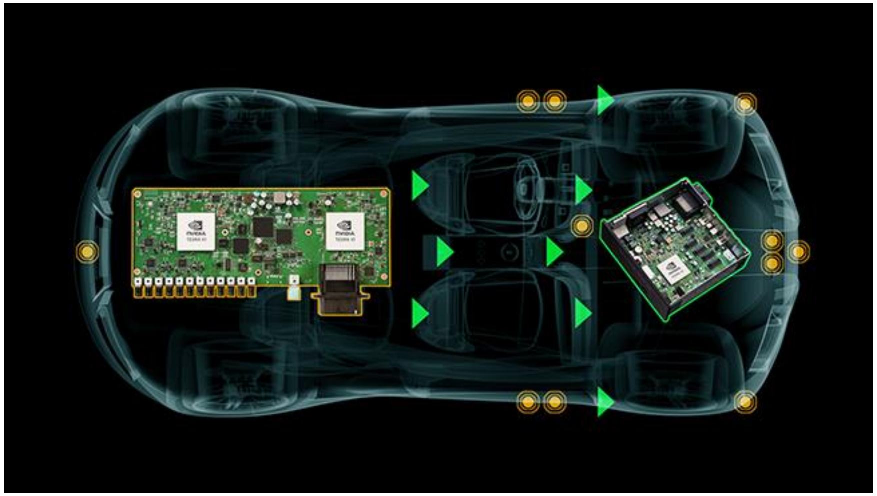
- Example training images for each orientation (Prof. H. Schneiderman)

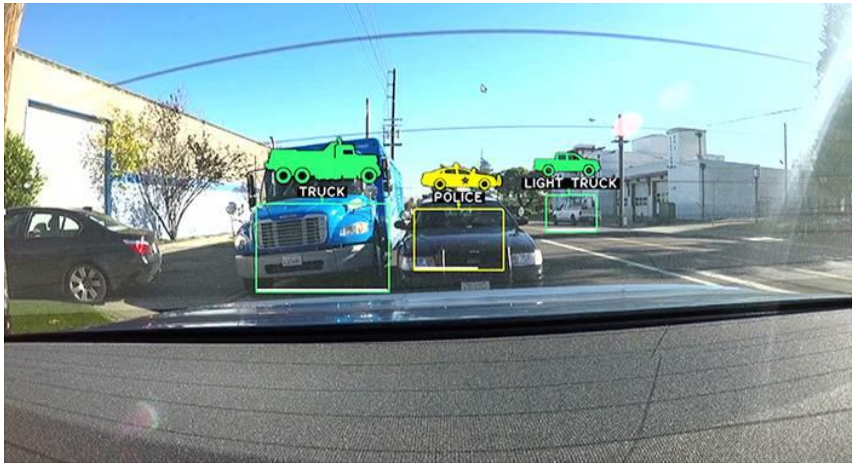
Pedestrian Detection

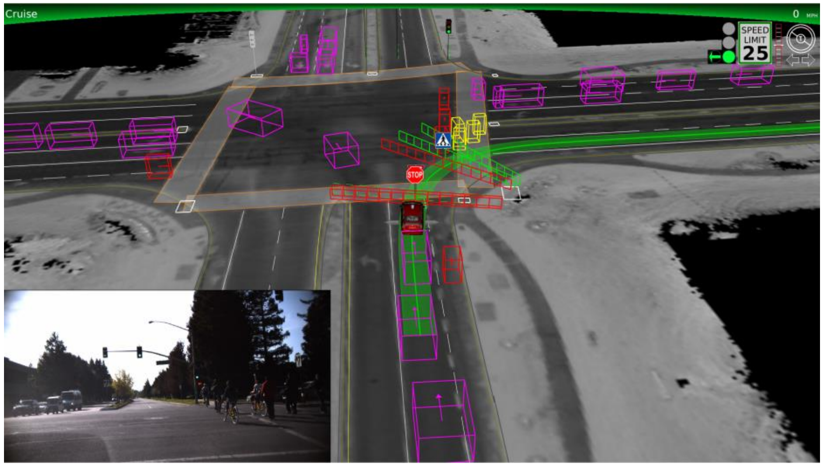


Autonomous

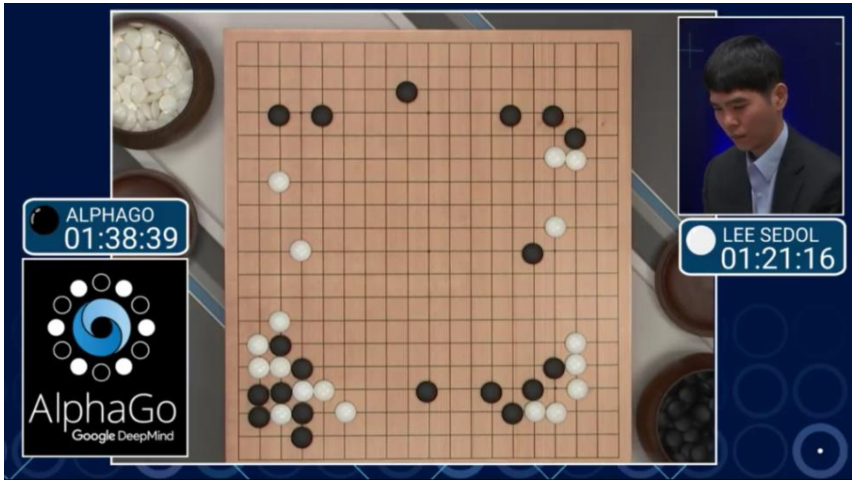

Google
self-driving car

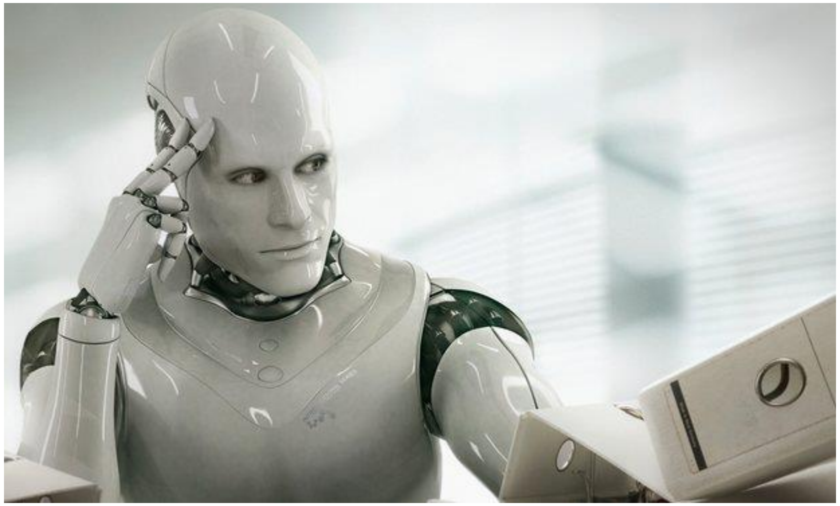


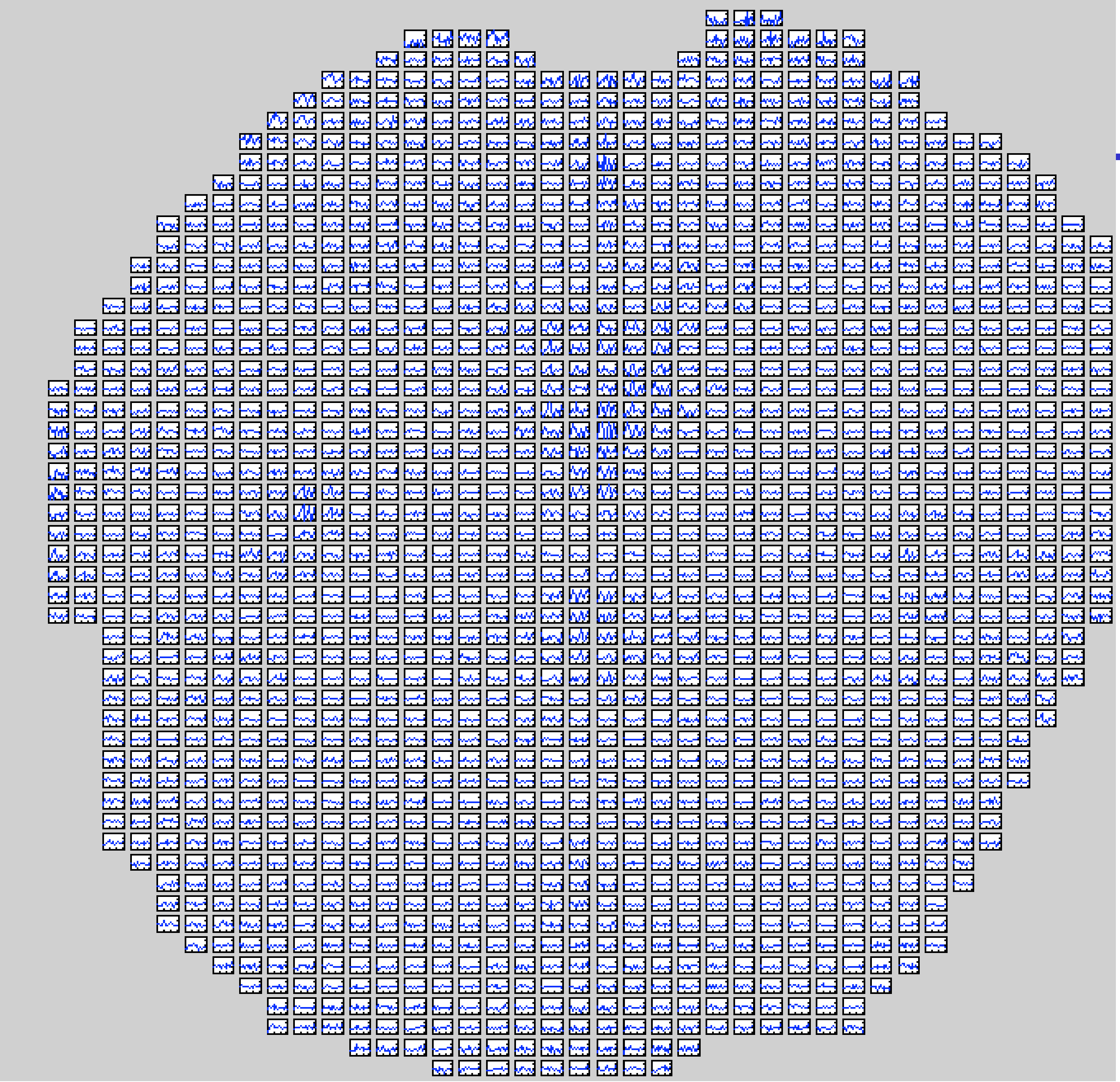

TRUCK
POLICE
LIGHT TRUCK


Cruise
SPEED LIMIT 25
STOP

AI and Robot







Reading a noun (vs verb)

Text Classification

Company home page

vs

Personal home page

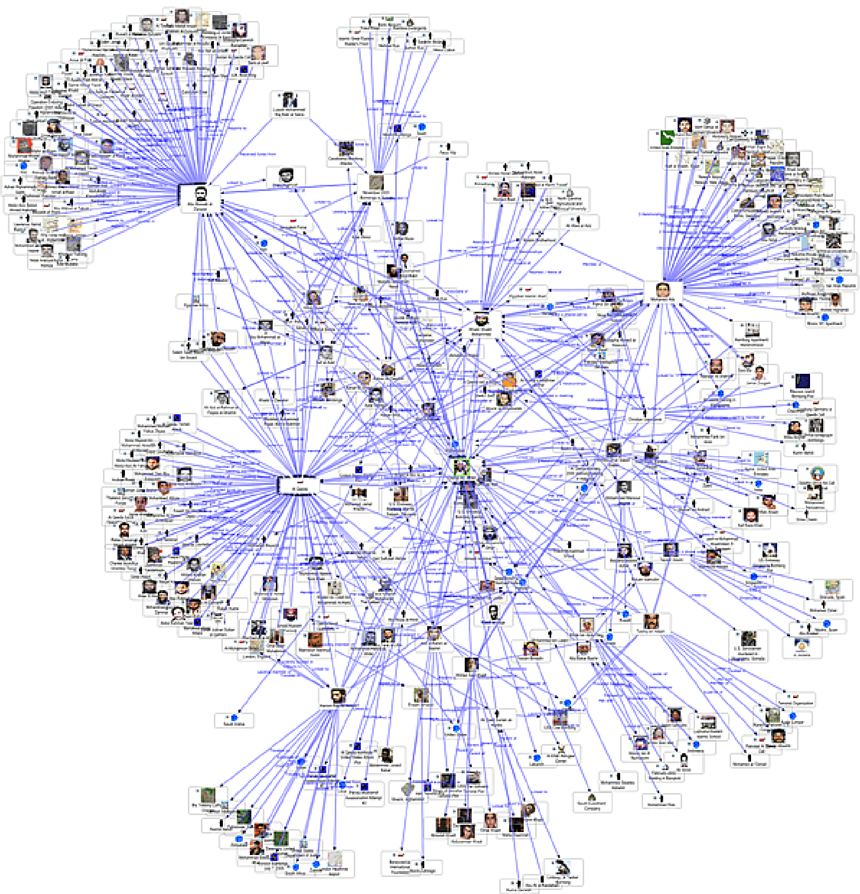
vs

University home page

vs

…

Social Network and Sentiment Analysis



"What people think?"

What others think has always been an important piece of information

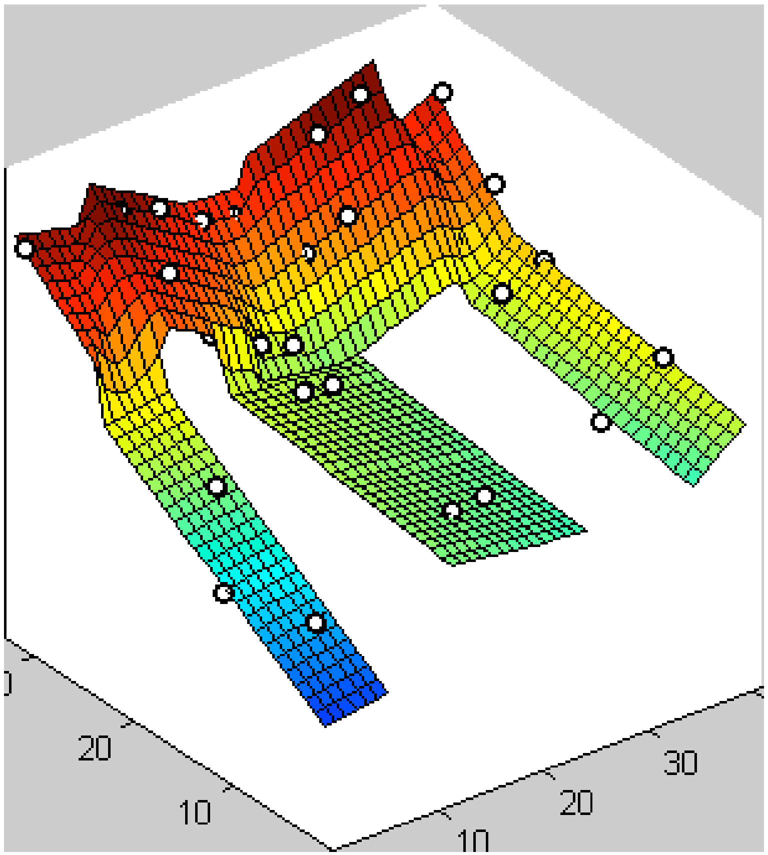
"Which car should I buy?"

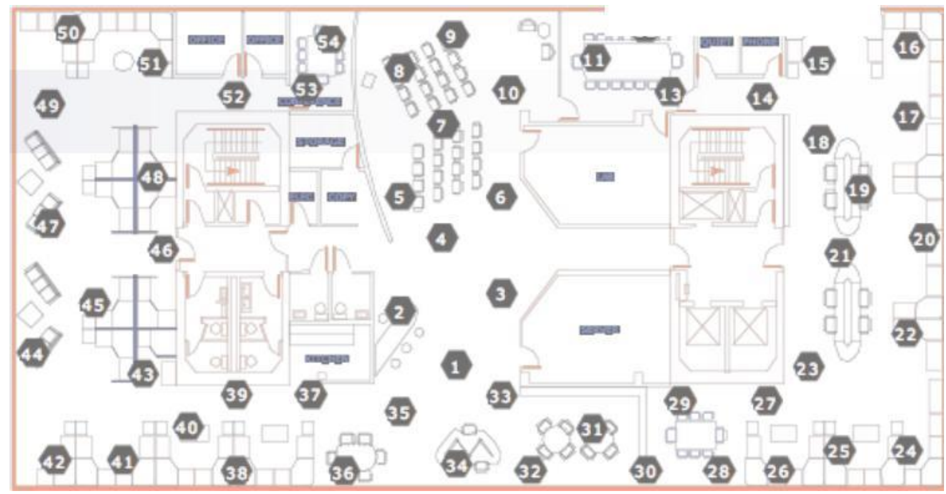
"Which schools should I apply to?"

"Which Professor to work for?"

"Whom should I vote for?"

Modeling Sensor Data



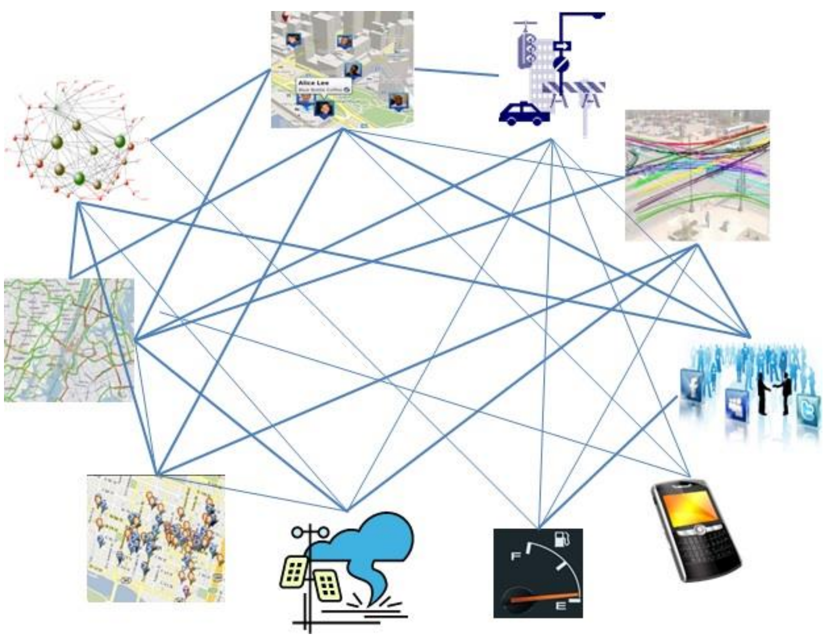


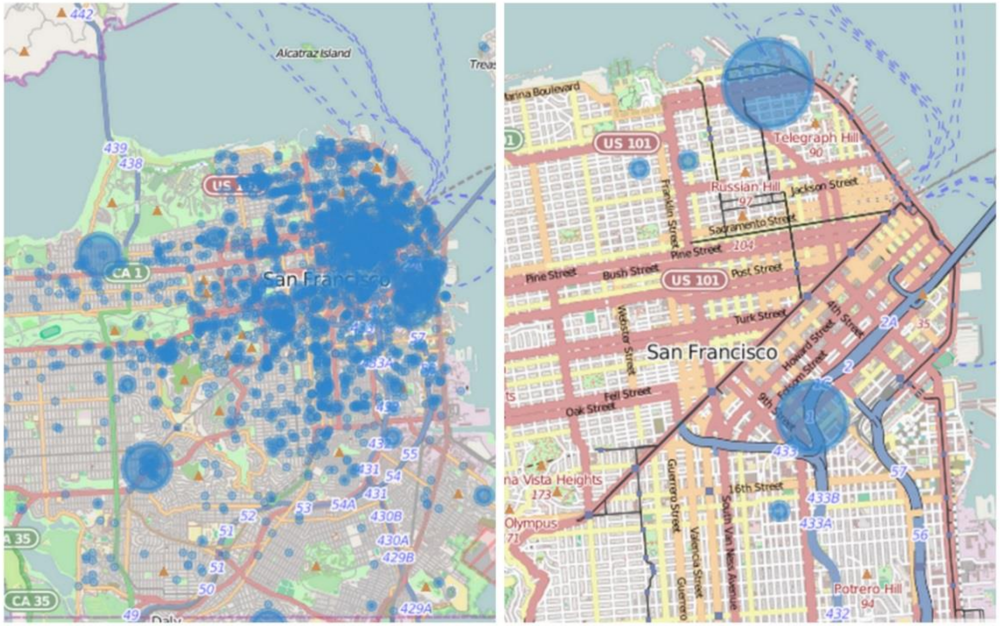
[Guestrin et al. '04]

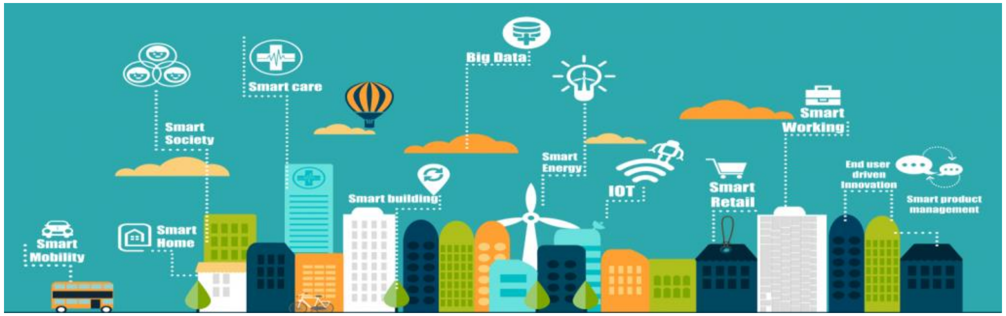
- Measure temperatures at some locations
- Predict temperatures throughout the environment

Urban Computing

- Uber, DiDi, etc.
- Cyber physical systems
- Internet of Things



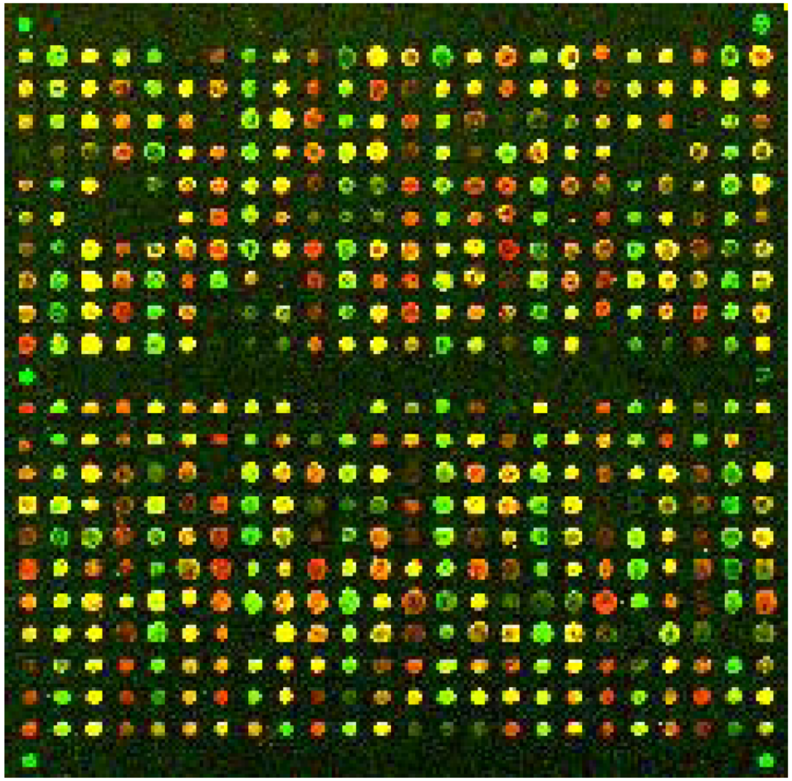


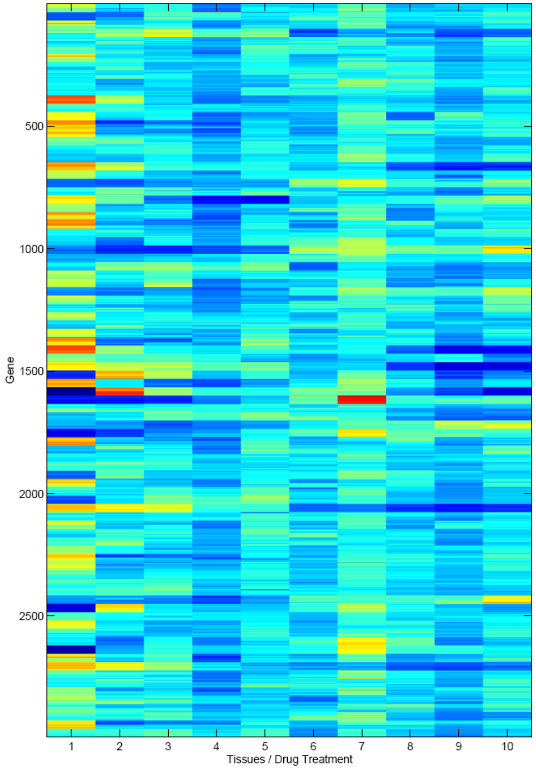


Financial Prediction

Bioinformatics

- E.g. modeling gene microarray data, protein structure prediction





Movie Recommendation Systems

- Challenge: to improve the accuracy of movie preference predictions
- Netflix $1m Prize. Competition started Oct 2, 2006 and awarded 2009.

More Applications

- **There are already a number of applications**
  - face, speech, handwritten character recognition
  - fraud detection (e.g., credit card)
  - recommender problems (e.g., which movies/products/etc you'd like)
  - annotation of biological sequences, molecules, or assays
  - market prediction (e.g., stock/house prices)
  - finding errors in computer programs, computer security
  - defense applications
  - etc

Types of Learning

- **Supervised learning**
  - you are given examples with correct labels and are asked to label new examples
- **Unsupervised learning**
  - you are given only unlabeled examples
- **Semi-supervised learning**
  - you are given both examples with correct labels and unlabeled examples
- **Active learning**
  - you are given an oracle, whom you can ask to label particular examples, then must label the others
- **Reinforcement learning**
  - you are given the overall performance (as opposed to labels to particular examples), not offered in this class
- **So many topics ……**

Supervised Learning

Data: $D = \{d_1, d_2, .., d_n\}$ **a set of *n* examples**

$d_i =< \mathbf{x}_i, y_i >$

$\mathbf{x}_i$ is input vector, and y is desired output (given by a teacher)

Objective: learn the mapping $f : X \to Y$

s.t. $y_i \approx f(x_i)$ for all $i = 1,.., n$

Two types of problems:

- **Regression:** X discrete or continuous $\longrightarrow$ Y is **continuous**
- **Classification:** X discrete or continuous $\longrightarrow$ Y is **discrete**

Ref: Milos Hauskrecht

Supervised Learning Examples

- **Regression:** Y is **continuous**

Debt/equity
Earnings
Future product orders

→ company stock price

- **Classification:** Y is **discrete**

```
# ######
      ##
     ##
  ######
      ###
#      ##
# ##   ##
  #######
```

→ Label "3"

Handwritten digit (array of 0,1s)

Ref: Milos Hauskrecht

Supervised Learning

- Example problem: face recognition

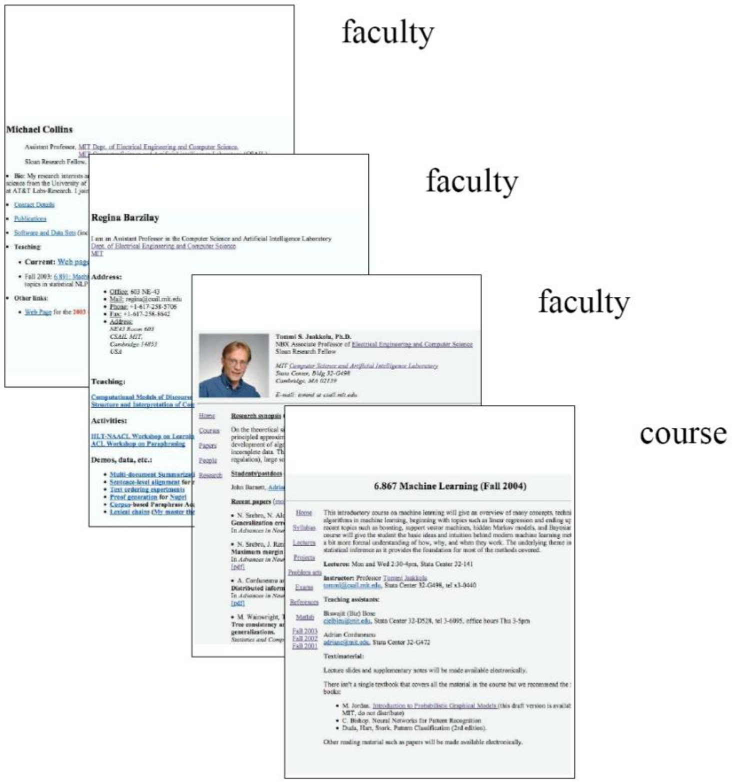
Training data: a collection of images and labels (names)

Evaluation criterion: correct labeling of new images

Ref: Tommi Jaakkola

Supervised Learning

- Example problem: text/document classification



A few labeled training documents (webpages)

Goal to label yet unseen documents

Ref: Tommi Jaakkola

Unsupervised Learning

- **Data:** $D = \{d_1, d_2, .., d_n\}$
  $d_i = \mathbf{x}_i$ vector of values
  No target value (output) y

- **Objective:**
  – learn relations between samples, components of samples

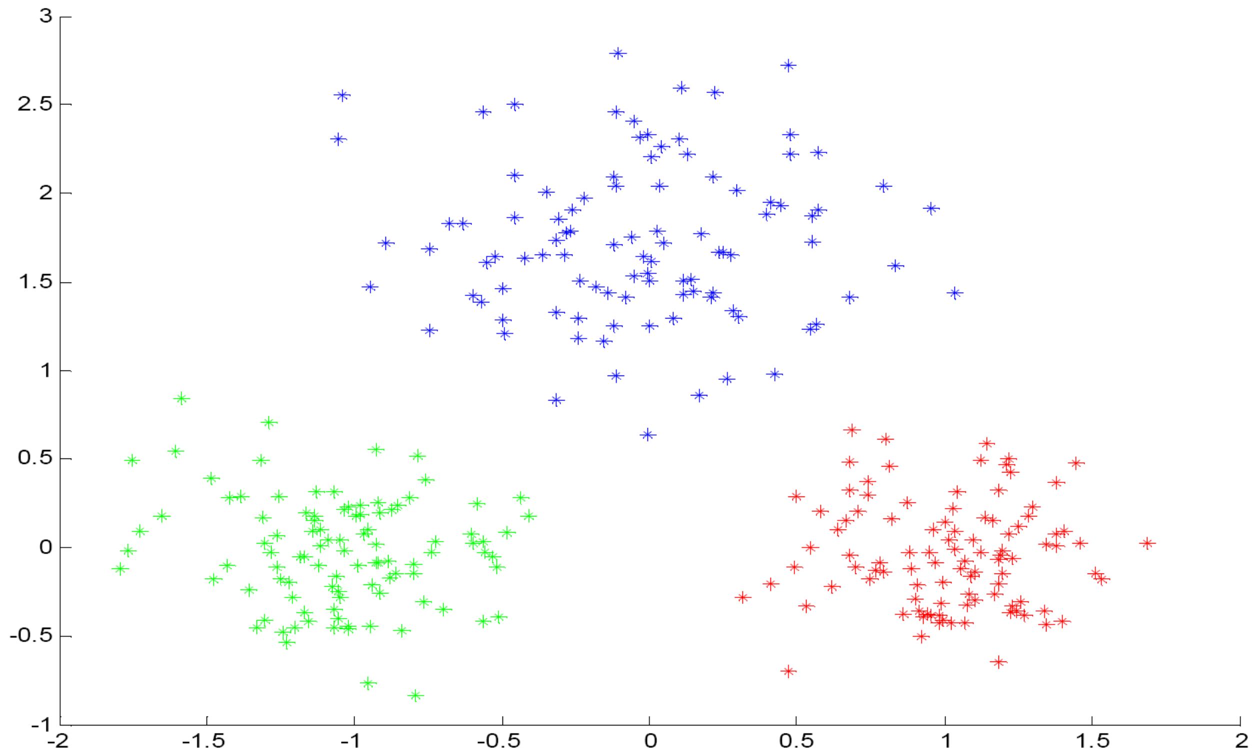
Types of problems:

- **Clustering**
  Group together "similar" examples, e.g. patient cases
- **Density estimation**
  – Model probabilistically the population of samples

Ref: Milos Hauskrecht

Unsupervised Learning Example

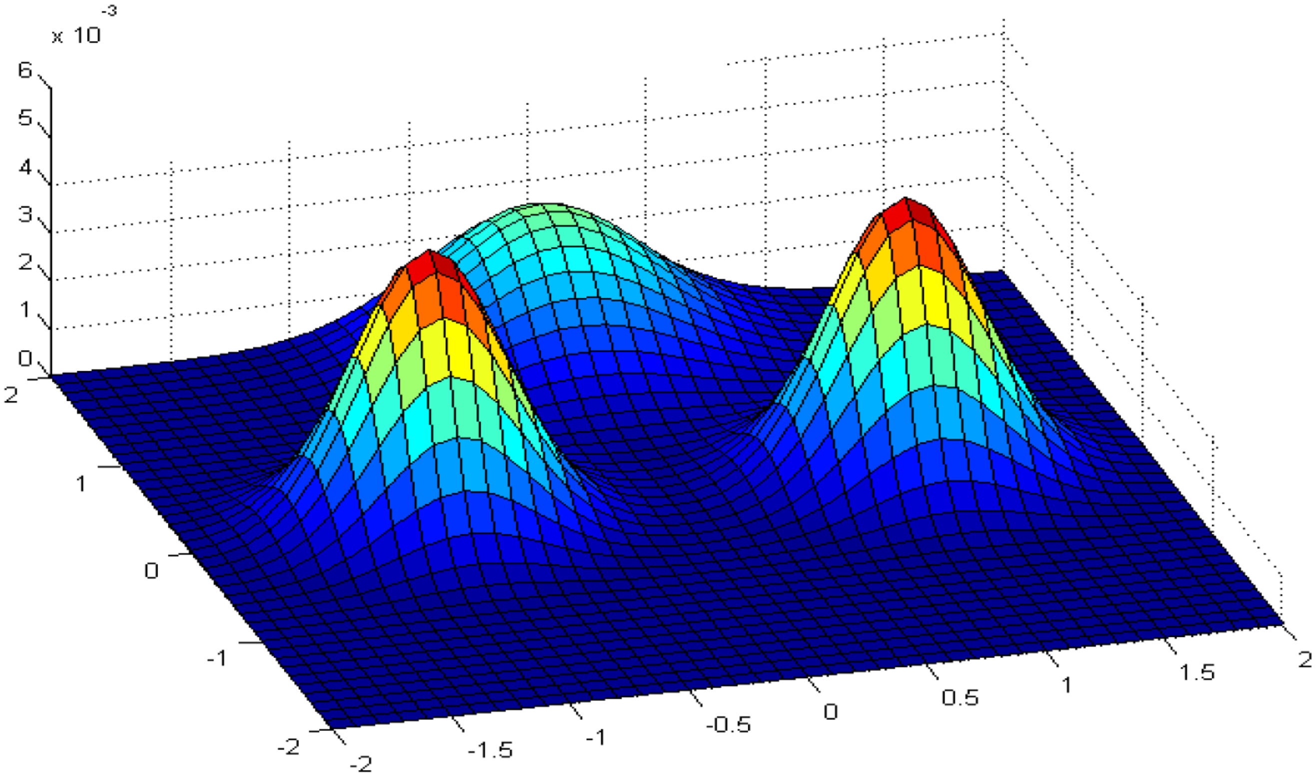
- **Density estimation.** We want to build the probability model of a population from which we draw samples $d_i = \mathbf{x}_i$



Ref: Milos Hauskrecht

Unsupervised Learning - Density Estimation

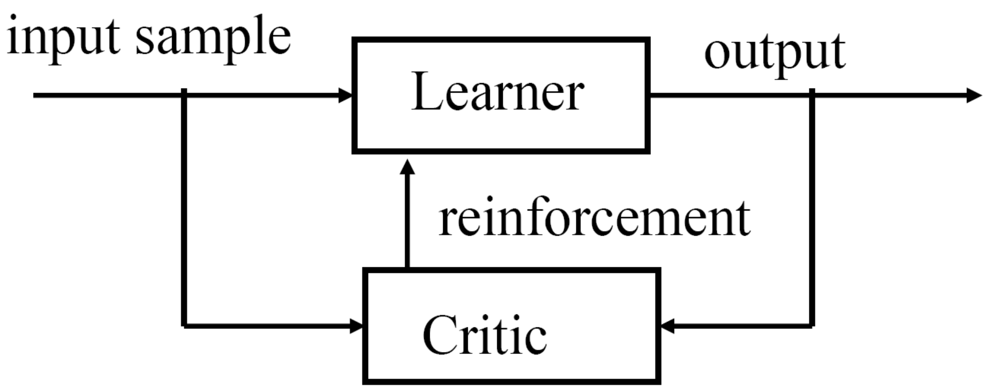
- A probability density of a point in the two dimensional space
  - Model used here: **Mixture of Gaussians**



Ref: Milos Hauskrecht

Reinforcement Learning

- We want to learn: $f : X \rightarrow Y$
- We see samples of **x** but not y
- Instead of y we get a feedback (reinforcement) from a **critic** about how good our output was



- The goal is to select outputs that lead to the best reinforcement

Ref: Milos Hauskrecht