MACHINE LEARNING FOR DATA MINING
LECTURE 1: INTRODUCTION

Siamak Sarmady (Urmia University of Technology)
Andrew Ng (Stanford University)
Pedro Domingos (Washington University)

A Few Quotes

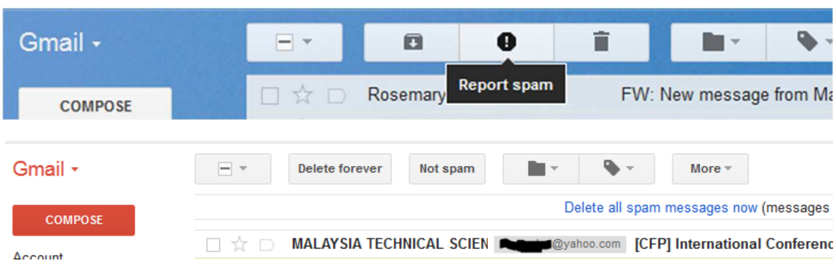
- "A breakthrough in machine learning would be worth ten Microsofts"(Bill Gates, Chairman, Microsoft)
- "Machine learning is the next Internet" (Tony Tether, Director, DARPA)
- "Machine learning is the hot new thing"(John Hennessy, President, Stanford)
- "Machine learning is going to result in a real revolution"(Greg Papadopoulos, Former CTO, Sun)
- "Machine learning today is one of the hottest aspects of computer science"(Steve Ballmer, CEO, Microsoft)

Sample Applications

- **Web search**
- **Finance** (e.g. stock price prediction)
- **E-commerce** (e.g. fraud detection)
- **Computational biology** (e.g. gene and DNA data analysis)
- **Space exploration** (control, image processing, …)
- **Robotics** (e.g. control, sensory data processing, decision making)
- **Information extraction** (and Data Mining)
- **Social networks** (e.g. friend and membership suggestions)
- **Software Debugging**
- **[Your favorite area]**

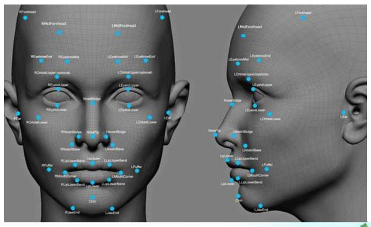
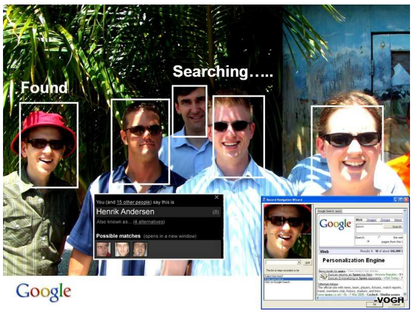
Applications

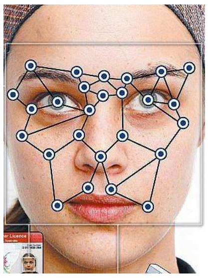
- We use machine learning everyday without knowing. Every time you use the anti-spam filters you are using machine learning…



Applications – Face Detection and Recognition

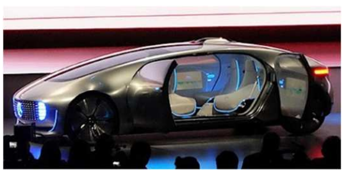
- Every time you use iPhoto, Google+, Facebook to automatically recognize you and your friends, you are using machine learning…

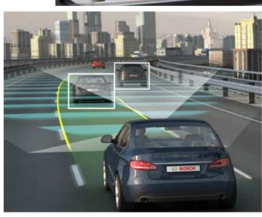







Applications – Self Driving Cars

- Self driving cars which will come to market this year are using machine learning… (Video)

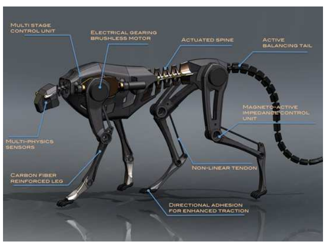



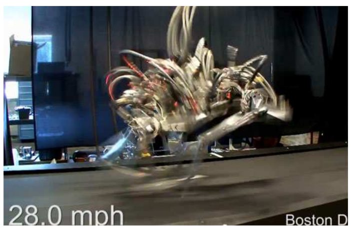



Applications – Self Driving Cars

Applications - Robots

- Robots are coming … military, surveillance, service… (See Boston Dynamics …) - Video





Applications – Robots (Boston Dynamics 2016)

Applications - UAVs

- Self learning, self driving aircrafts…. Self piloted drones… Drone control in the absence of satellite communications … (Video)

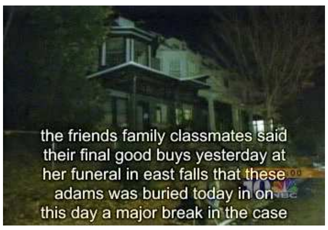


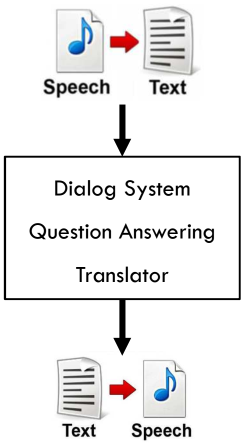



Applications– Stanford Self-Learning Helicopter (reinforcement learning)

Natural Language Processing (NLP)

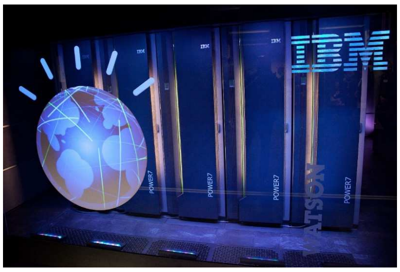
- Speech technologies (e.g. Cortana, Siri)
  - Automatic speech recognition (ASR) or Speech to Text (STT)
  - Text-to-speech synthesis (TTS)
  - Dialog systems
    - Question answering



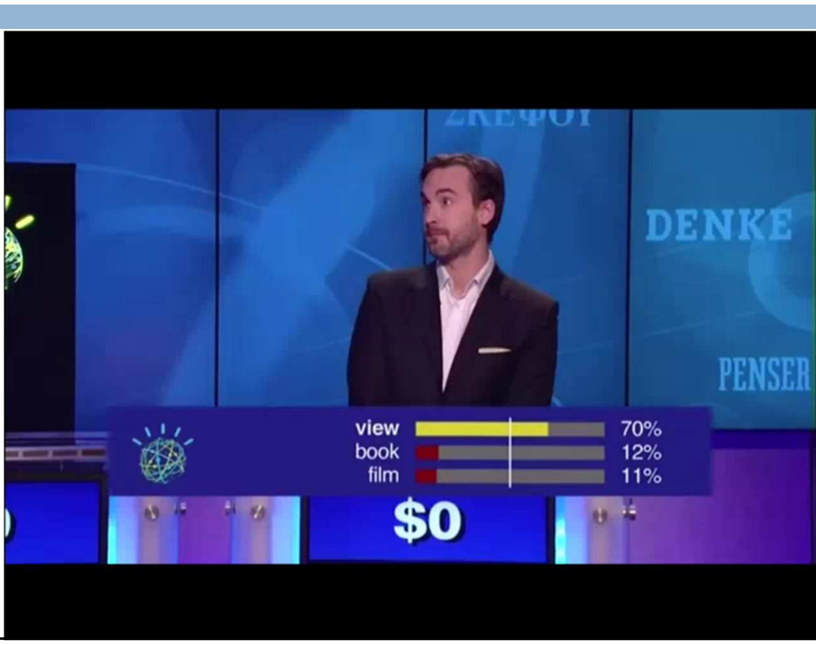


Natural Language Processing (NLP)

- Question Answering, Dialogue systems …
  - Mostly use Deep Neural Networks (Video)

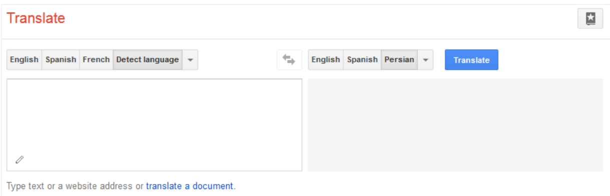


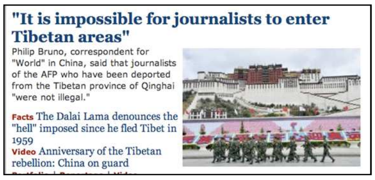
Applications - Watson (IBM)

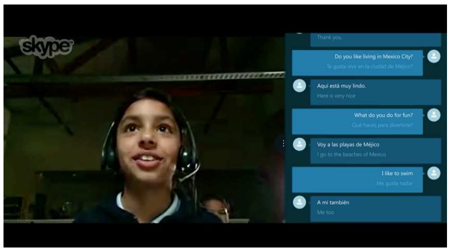


Natural Language Processing (NLP)

- Other NLP technologies
  - Machine translation
  - Web search
  - Text classification, spam filtering and etc.
  - Text correction (dictation, grammar)







Machine Learning

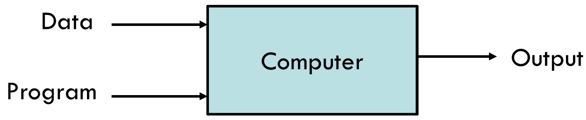
- Grew out of work in AI field.
  - When AI based on search and logic did not become successful enough, some researchers started modern AI using statistical and learning methods.

- Brought new capabilities to computers, such as:
  - Mining of large datasets from growth of automation/web
    - Web click data (advertise, better services), medical records (obtain medical knowledge), biology, engineering
  - Applications that cannot be programed by hand (we don't know how to do that):
    - Autonomous helicopter, handwriting recognition (mail), most of Natural Language Processing (NLP), Computer Vision.
  - Self-customizing programs
    - E.g., Amazon, Netflix product recommendations
  - Understanding human learning (brain, real AI).

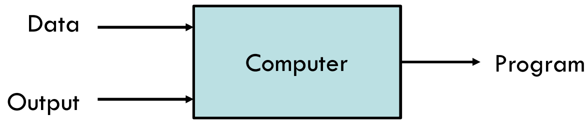
Machine Learning (vs. Normal Programs and Classic AI)

- Writing software is the bottleneck of building computer systems
- So, how about getting computers to program themselves
- Give the data to computer, Let it create the program itself
- This is in fact automating automation

Traditional Programming:



Machine Learning:



Classic AI vs. Machine Learning vs. Traditional Programming

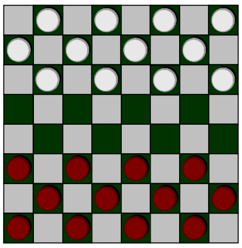
- **Traditional Programming vs. Machine Learning**
  - In some situations we don't know how to design an algorithm to solve the problem (e.g. face detection)
  - Traditional programming methods are incapable when it comes to very complicated scenarios… learning is the way forward
- **Classic AI vs. Machine Learning**
  - Classic AI (i.e. search, logic and symbolic methods) has not been very successful in building General AI
  - Classic AI does not work well in dynamic, uncertain and non-deterministic environments
  - The best bet is to mimic natural AI and the way it works (aka. Neural networks)

Machine Learning definition

- **Arthur Samuel's (1959) definition of machine learning:** Field of study that gives computers the ability to learn without being explicitly programmed.

  He created a checkers program that played the game tens of thousands of times against itself and learned what positions are good or bad. It was then able to play much better than an average human (**reinforcement** learning).

- **Tom Mitchell's (1998) description of Learning:**
  - Assume we have a "task T".
  - A program gathers "experience E" by doing T (or by watching someone doing it)
  - The performance is measured by a "performance measure P"
  - If "performance measure P" improves by experience E, then the program is learning.



Checkers, Chess, …

Quiz

- "A computer program is said to learn from experience E with respect to some task T and some performance measure P, if its performance on T, as measured by P, improves with experience E."

- Suppose your email program watches which emails you do or do not mark as spam, and based on that learns how to better filter spam. Which items match T, P, and E?
  - Classifying emails as spam or not spam.
  - Watching you label emails as spam or not spam.
  - The number (or fraction) of emails correctly classified as spam/not spam.
  - None of the above—this is not a machine learning problem.

Types of Learning

- **Supervised/inductive learning (classification or regression)**
  - Training data includes desired outputs. We provide a series of "input->output" pairs. The algorithm learns from them. We then provide inputs and the trained algorithm tries to guess an output.
- **Unsupervised learning (clustering)**
  - Training data does not include desired outputs. We only adjust the algorithm parameters in a way that the inputs are clustered into separate groups based on specific similarities
- **Semi-supervised learning**
  - Training data for clustering includes a few desired outputs (labels)
- **Reinforcement learning**
  - We only provide a performance measure. The algorithm randomly tries different things (looking at their performance). It will then repeat those actions that produce better results.
- **Recommender system**
  - Looks into the selections we make, it tries to select the same way (can be done using clustering, so could be an application of above)

What We'll Cover

- Covered
  - Intro
  - Naive Bayes
  - Neural networks
  - SVM
  - Decision Trees
  - kNN
  - Regressions
  - Unsupervised learning, clustering and dimensionality reduction
  - PCA
- Selected
  - Feature Scaling
  - Text Learning
  - Feature Selection
  - Model ensembles
  - Validation
  - Evaluation Metrics
  - Recommender systems
  - Large-scale machine learning
  - Practical advice for applying learning algorithms

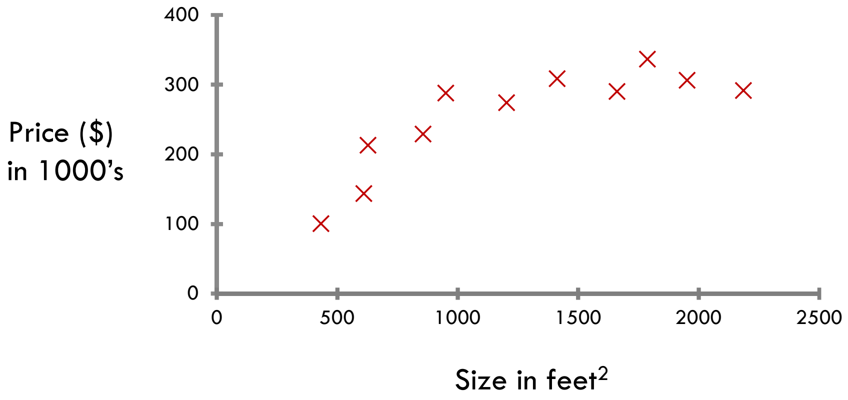
Supervised Learning

Inductive (Supervised) Learning

- If we use examples of a function (i.e. some x and F(x) values) to build a function (or model) that can predict F(x) for new values x … then we have performed supervised learning.
  - **Discrete F(x) :** Classification
  - **Continuous F(x) :** Regression
  - **F(X) = Probability(X) :** Probability estimation
    - in fact a regression with output value in the range of [0,1]

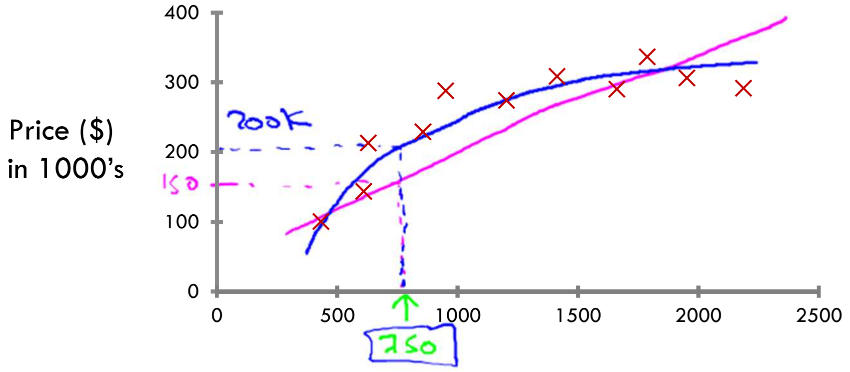
Housing price prediction

- Having this data, assume you want to predict the price of your friend's 750 ft^2 house for him… How a learning algorithm help you?



Housing price prediction

- The learning algorithm can use a linear fit
- Or better, it can fit a more accurate curve (e.g. a quadratic function) and do a better prediction…



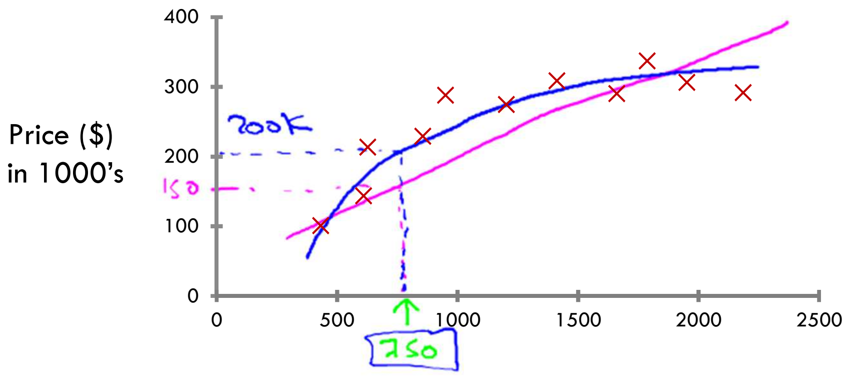
Housing price prediction

Supervised Learning:

Examples of "right answers" given

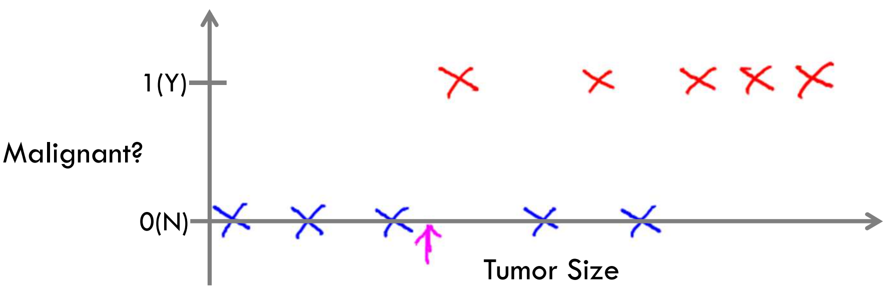
Regression:

Predict continuous valued output (price)

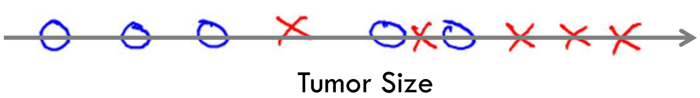


Breast cancer (malignant, benign)

- Assume you look into medical records and you want to predict whether a Tumor with the specified size is malignant or benign based on its size…



Alternative presentation:



Classification:

Discrete valued output (0 or 1)

Not necessarily limited to two classes:

0: Benign

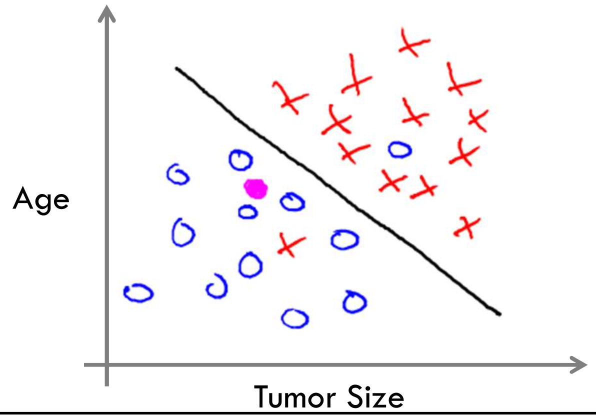
1: Type1

2: Type2

3: Type3

Multiple Parameters

- In previous example, the prediction was based on only one parameter (tumor size). Let's assume we have extracted two parameters from the medical records. If the new parameter is relevant, it might help in better classification.
- The algorithm again tries to find a line (or barrier) that separates the two classes (malignant, benign).



Other potential parameters:

- Clump Thickness
- Uniformity of Cell Size
- Uniformity of Cell Shape

…

How supervised learning typically works

- We start by choosing a model-class: $y = f(\mathbf{x}; \mathbf{W})$
  - A model-class, f, is a way of using some numerical parameters, **W**, to map each input vector, **x**, into a predicted output y.
- Learning usually means adjusting the parameters to reduce the difference between the target output, t, on each training case and the actual output, y, produced by the model.
  - We use numerical measures to minimize the difference between predicted output and the actual output
    - For regression, we will see later that $\frac{1}{2}(y-t)^2$ is often a suitable measure.
    - For classification there are other measures that are generally more sensible (they also work better).

Recap

- In this course we will be looking at supervised learning methods. The idea is that in our training data set, we are going to give the algorithm some correct answers.

- The algorithm will learn from the training set and it will generalize what it has learned to the questions it has not seen.

- Learning is done by adjusting the model parameters (e.g. using optimization methods).

- It will guess the answers for new questions, based on what it has learned in past.

- **Classification:** discrete
- **Regression:** continuous

Quiz

- You're running a company, and you want to develop learning algorithms to address each of two problems.
- **Problem 1:** You have a large inventory of identical items. You want to predict how many of these items will sell over the next 3 months.
- **Problem 2:** You'd like software to examine individual customer accounts, and for each account decide if it has been hacked/compromised.

- Should you treat these as classification or as regression problems?
  - Treat both as classification problems.
  - Treat problem 1 as a classification problem, problem 2 as a regression problem.
  - Treat problem 1 as a regression problem, problem 2 as a classification problem.
  - Treat both as regression problems.

Methods and Representations of Classification

- Naïve Bayes
- Neural networks
- Decision trees
- Support vector machines
- kNN
- Sets of rules / Logic programs
- Graphical models (Bayes/Markov nets)
- Model ensembles
- …

Evaluation Measures

- Accuracy
- Squared error
- Precision and recall
- Likelihood
- Posterior probability
- Cost / Utility
- Margin
- Entropy
- K-L divergence
- …

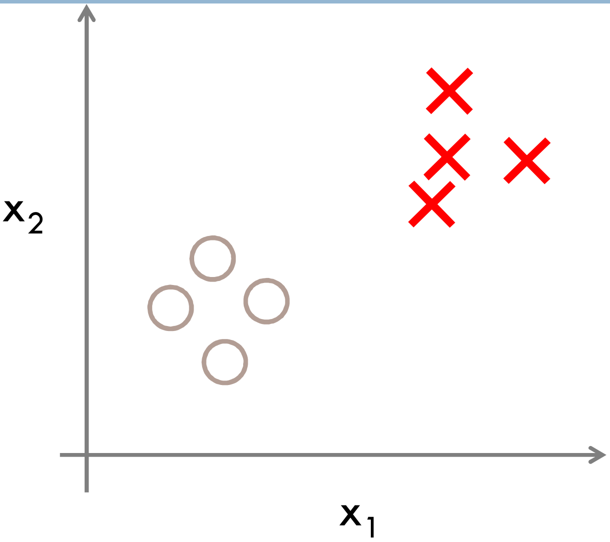
Optimization Methods

- Optimization methods:
  - Combinatorial optimization
    - Greedy search
  - Convex optimization
    - Gradient descent
  - Constrained optimization
    - Linear programming

Unsupervised Learning

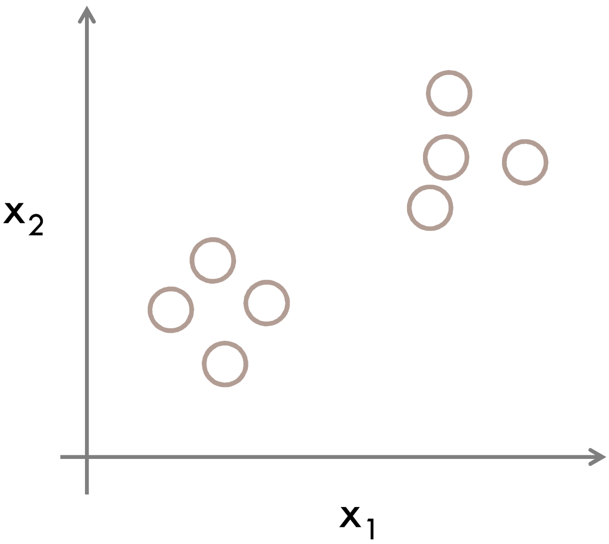
Supervised Learning

- In supervised learning, the training data set provided the correct "labels" (i.e. classes) for individual data (e.g. O and X here).
- What if we don't have labels for the data points?



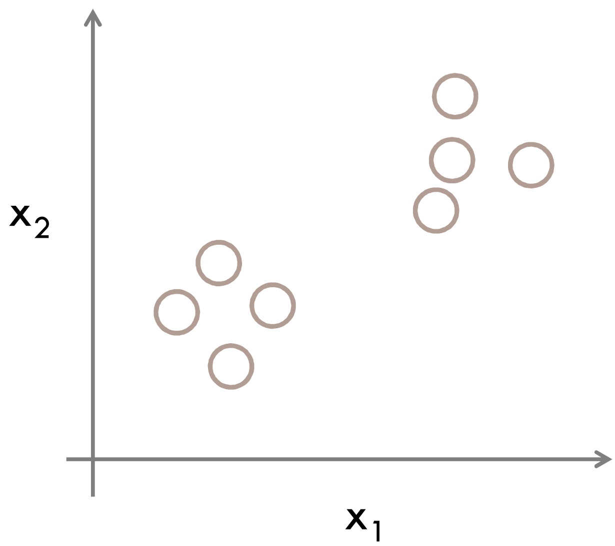
Unsupervised Learning

- In these kind of problems, we are given some data but we don't know how different they are and there are no labeled examples
- We are asked to find a structure in data.



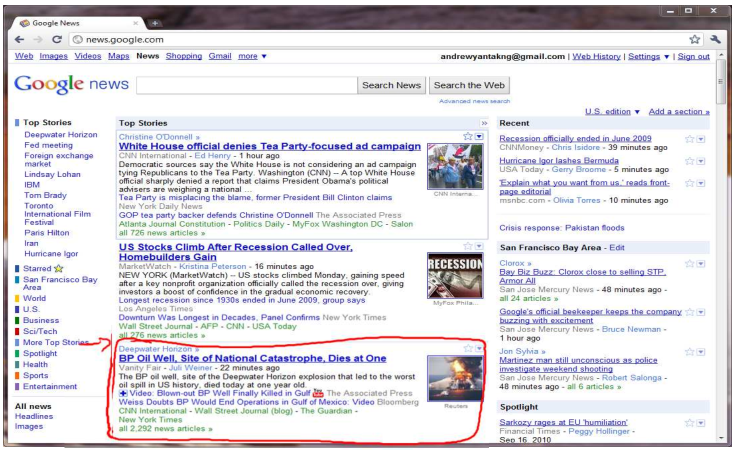
Unsupervised Learning

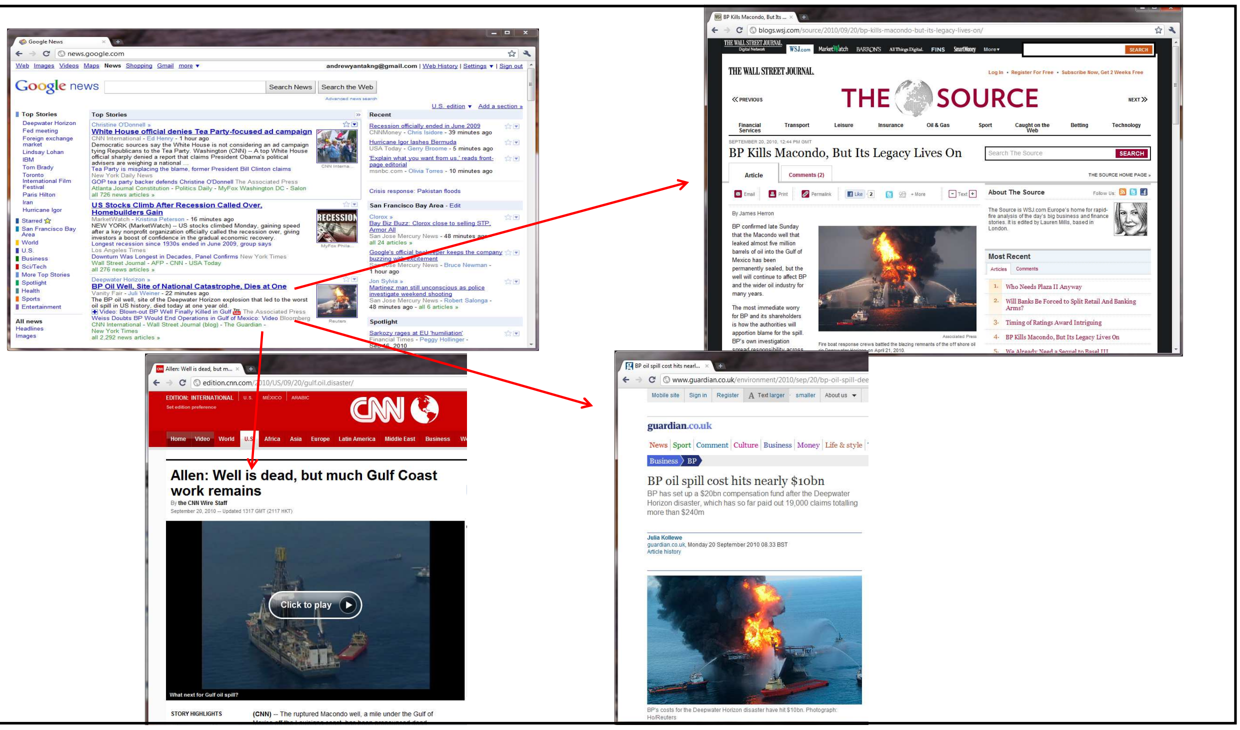
- The learning algorithm might decide that in this data there are two clusters of data… this is called clustering



Unsupervised Learning - Applications

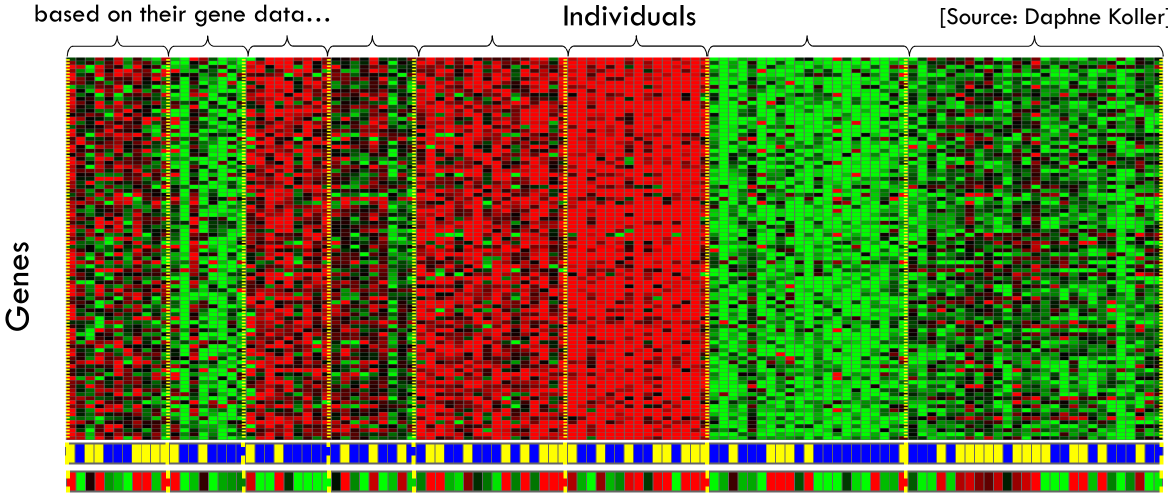
- One example of clustering is used in Google news. News items of the same topic are clustered into separate subjects (headlines). No label or supervision is provided … It just recognizes clusters of news… (using words in the article)





Clustering - Applications

- Here is another example of clustering. We have gathered gene information of different individuals. We want to see whether different people have specific genes… and then we want to divide people into different categories or types … note that we don't know what exactly are those genes… we just cluster the individuals based on their gene data… [Source: Daphne Koller]



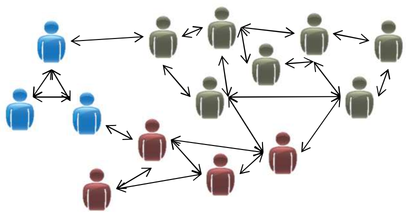
Why called unsupervised

- Because we don't provide the right answer (labels) to the algorithm, and the algorithm finds similarities in them automatically, it is called unsupervised learning….

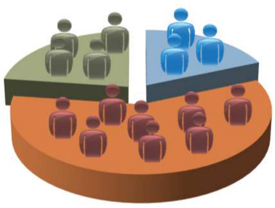
- It is the given examples that supervise the learning in supervised methods.

Applications

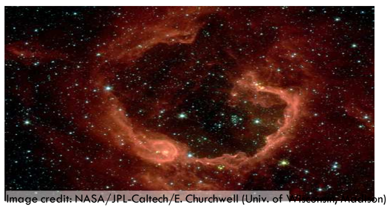
Organize jobs on clusters



Social network analysis



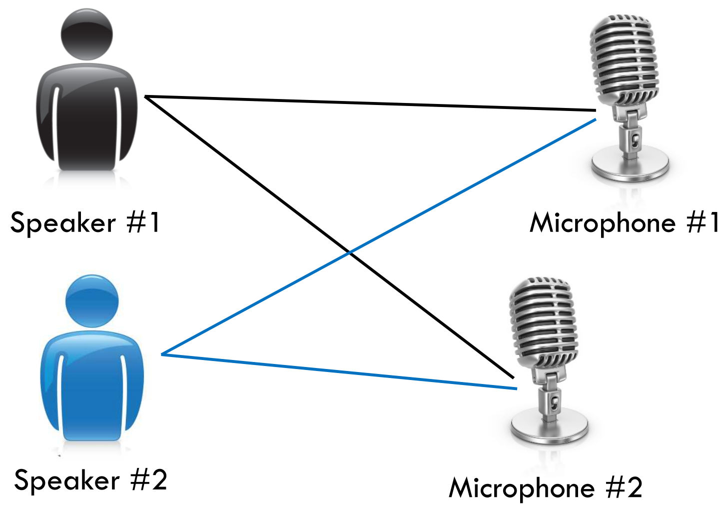
Market segmentation



Astronomical data analysis

Applications- Voice Filtering - Party problem

- In a busy party, the sounds are overlapped and we can hardly recognize what a specific person is saying.
- Using unsupervised learning we can separate the voices of different people or devices…
- We use two microphones at different places… so each records a different combination of voices… so let's say person #1 is more near to mic #1 and person #2 is nearer to mic #2.



Applications- Voice Filtering

Microphone #1:	Output #1:
Microphone #2:	Output #2:

Microphone #1:	Output #1:
Microphone #2:	Output #2:
Microphone #3:	Output #3:

[Audio clips courtesy of Te-Won Lee.]

Applications- Voice Filtering - algorithm

The whole separation of the voices can be done with a single line of MATLAB or OCTAVE code… that uses SVD function (Singular value decomposition).

```
[W,s,v] = svd((repmat(sum(x.*x,1),size(x,1),1).*x)*x');
```

Most learning problems can be solved with a few lines of code in these environments (the libraries could however be large).

We normally do the prototype program in these environments and then convert it into faster Java or C++ code…

If you are interested in details of the algorithm, please see the following source:
[Sam Roweis, Yair Weiss & Eero Simoncelli]

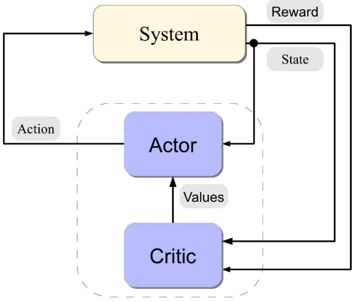
Quiz

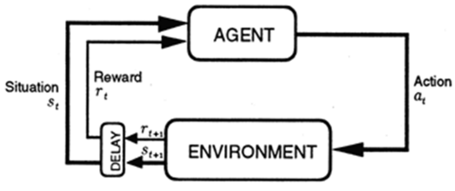
Of the following examples, which would you address using an <u>unsupervised</u> learning algorithm? (Check all that apply.)

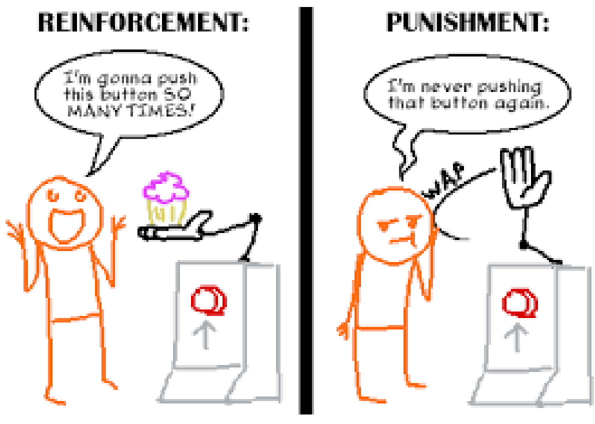
- ☐ Given email labeled as spam/not spam, learn a spam filter.
- ☐ Given a set of news articles found on the web, group them into set of articles about the same story.
- ☐ Given a database of customer data, automatically discover market segments and group customers into different market segments.
- ☐ Given a dataset of patients diagnosed as either having diabetes or not, learn to determine new patients as having diabetes or not.

Reinforcement Learning

Reinforcement learning







Figures: Credits belong to respective owners

Reinforcement learning

- In reinforcement learning, the output is an action or sequence of actions and the only supervisory signal is an occasional scalar reward.
  - The goal in selecting each action is to maximize the expected sum of the future rewards.
  - We usually use a discount factor for delayed rewards so that we don't have to look too far into the future.
- Reinforcement learning is difficult:
  - The rewards are typically delayed so its hard to know where we went wrong (or right).
  - A scalar reward does not supply much information.
- This course cannot cover everything and reinforcement learning is one of the important topics we will not cover.