Pattern Information Processing
(パターン情報処理)

Masashi Sugiyama
(Department of Computer Science)
杉山　将（計算工学専攻）

Contact:　W8E-505
sugi@cs.titech.ac.jp
http://sugiyama-www.cs.titech.ac.jp/~sugi

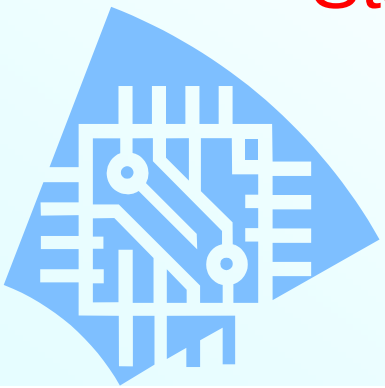
Contents of This Lecture (1)

- Syllabus (what I will provide in this course):

  Inferring an underlying input-output dependency from input and output examples is called supervised learning.

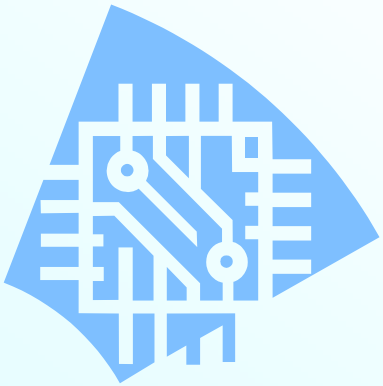
  This course focuses on a statistical approach to supervised learning and introduces its basic concepts as well as state-of-the-art techniques.

Statistical machine learning

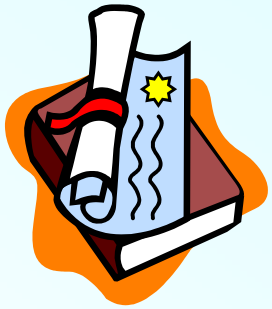


Contents of This Lecture (2)

- What you are expected to learn in this course:
  - How to use supervised learning methods
  - Ideas behind the methods
  - Novel research topics in supervised learning
  - Something useful in your own research/life



Grading

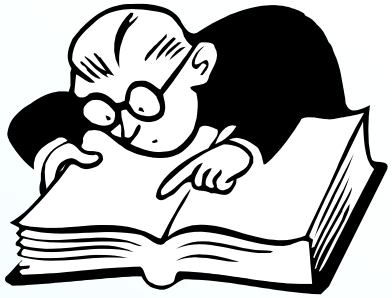


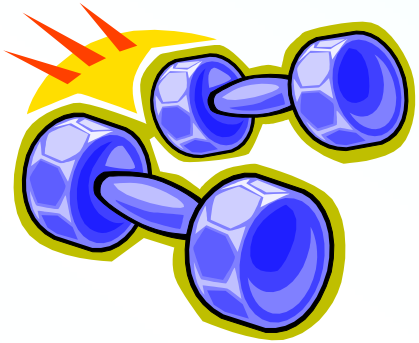
- Small reports
  - Almost every week
  - Deadline: next class
- Mini conference on supervised learning (final day)
  - Apply supervised learning techniques to your own data sets and analyze them!
- Final reports on the above issue

Brief Overview of the Course (1)

3 types of learning

- Supervised learning
- Unsupervised learning
- Reinforcement learning





Brief Overview of the Course (2)

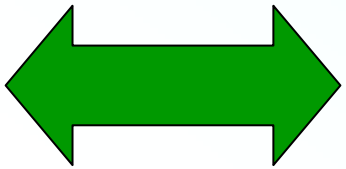
- Topics in supervised learning
  - Active learning
  - Model selection
  - Learning method

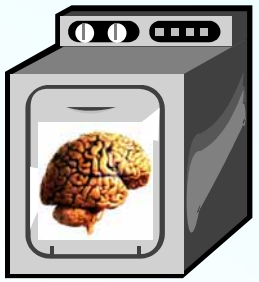
Textbook

- Handouts are provided if necessary.
- Pointers to related articles will be provided.

- I suppose you have elementary statistics and linear algebra. If not, please study them by yourself!

3 Topics in Learning Research



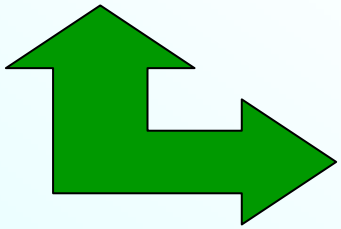


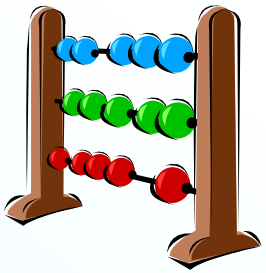
Understanding the brain

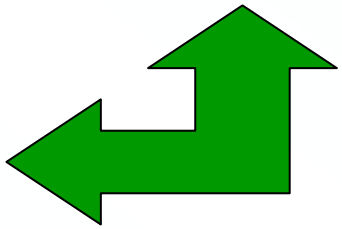
(Physiology, psychology, neuroscience)

Developing learning machines

(Computer and electronic engineering)







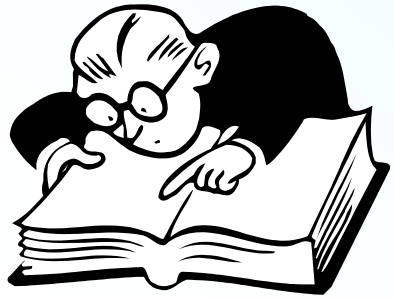
Mathematically clarifying mechanism of learning

(Computer and information science)

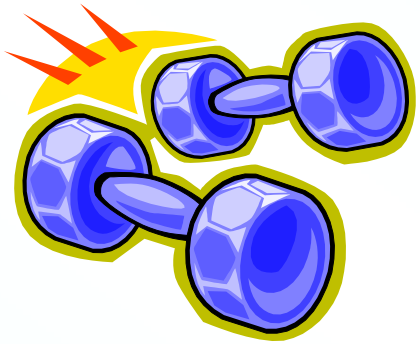
Three Types of Learning

- Supervised learning
  (This course!)

- Unsupervised learning

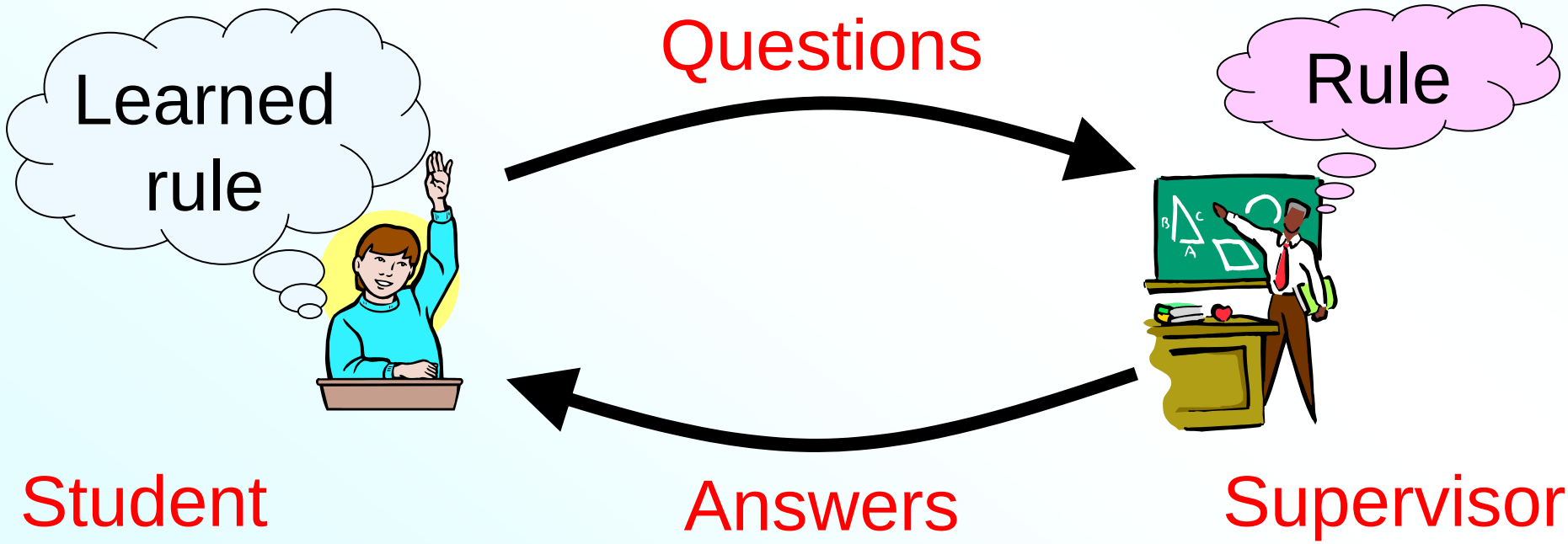


- Reinforcement learning



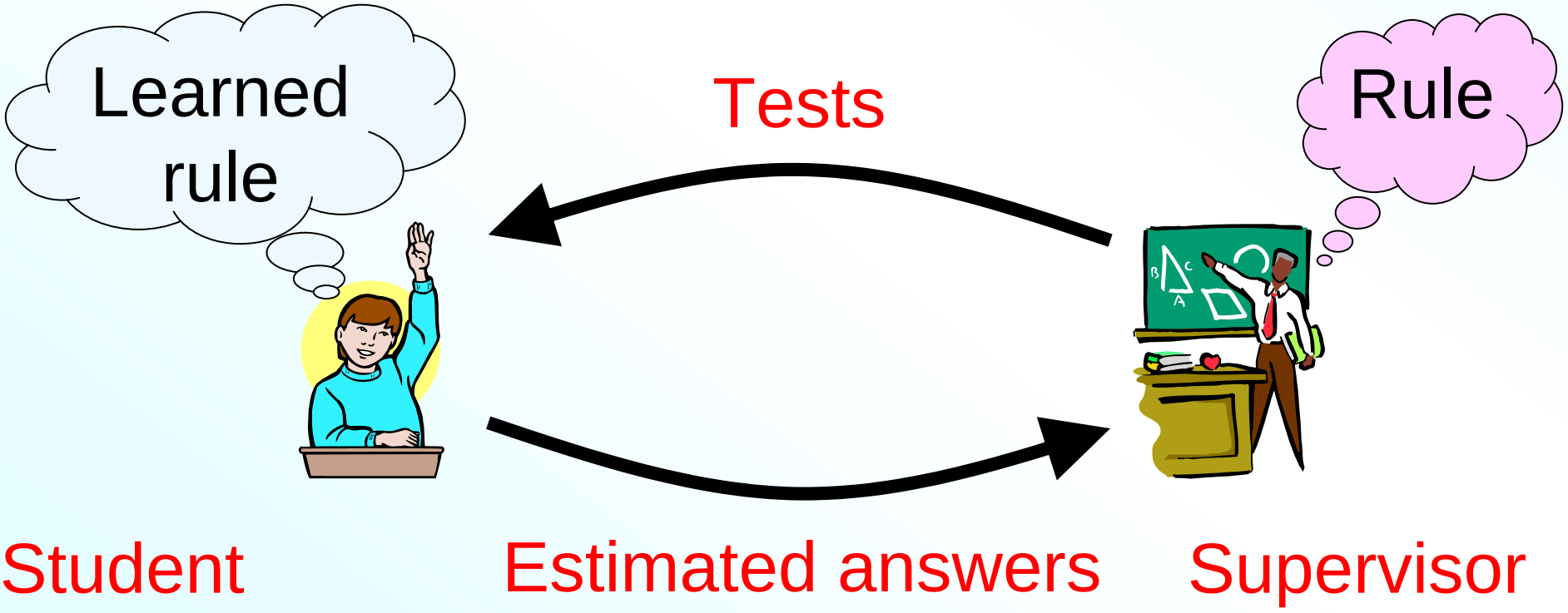
What Is Supervised Learning?

- The goal of supervised learning is to estimate an unknown input-output rule.
- You are allowed to ask questions to a supervisor ("oracle") who knows the rule.
- The supervisor answers your questions using the rule.



Generalization Capability

- Training examples: pairs of questions and answers.
- If the underlying rule can be successfully estimated, we can answer to the questions that we have never taught.
- Such an ability is called the generalization capability.



Example 1

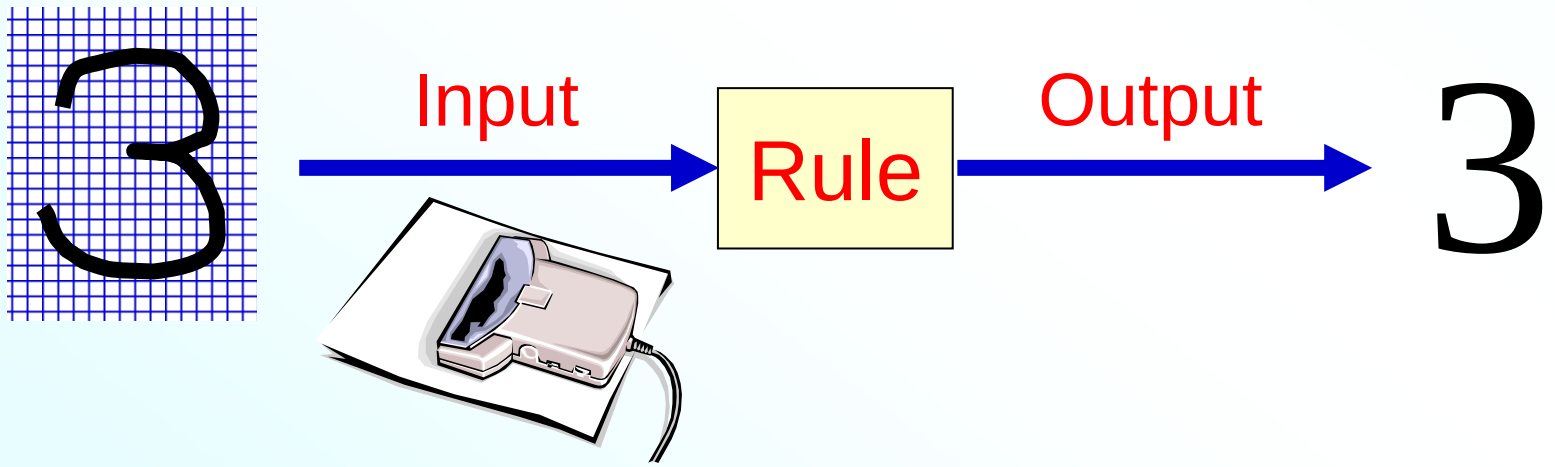
- Hand-written number recognition

  We want to recognize the scanned hand-written characters.

MNIST data from LeCun et al., Proc. IEEE, 1998

Example 1

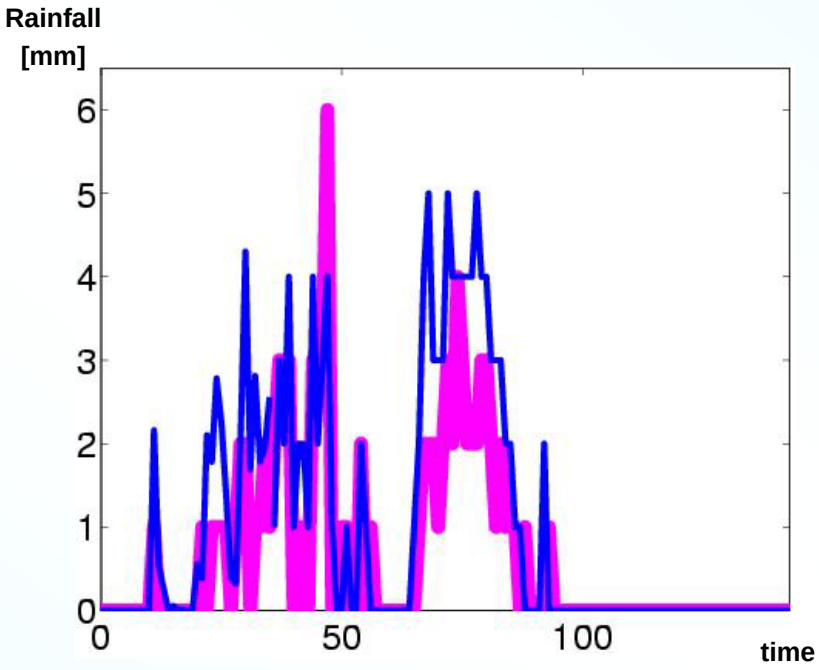
- Training examples consist of
  { (hand-written number, its recognition result) }.
- If underlying input-output rule is successfully learned, unlearned hand-written numbers can be recognized.



Example 2

- Rainfall Estimation

  Using the past rainfall and weather radar data, we want to estimate the rainfall tomorrow.



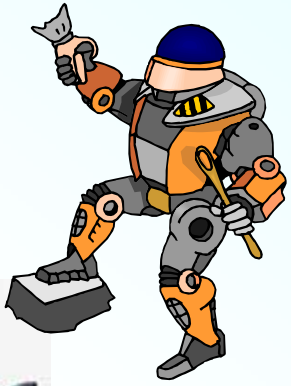
Weather radar data from IEICE Precipitation Estimation Contest 2001

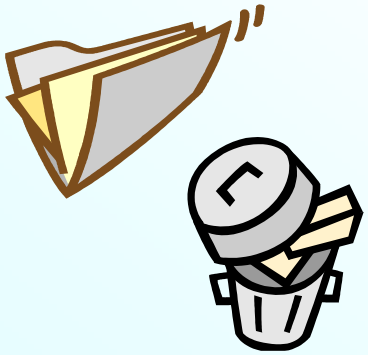
Example 2

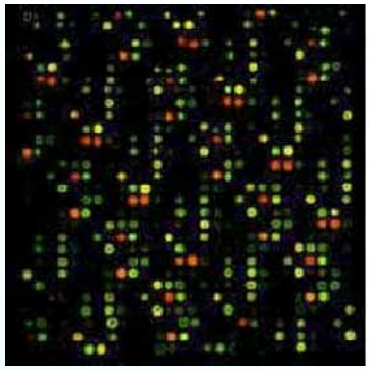
- Training examples are
  {(past rainfall and radar data,
  rainfall the next day)}
- If the rule is successfully learned, we can estimate the future rainfall by using the past rainfall and radar data.

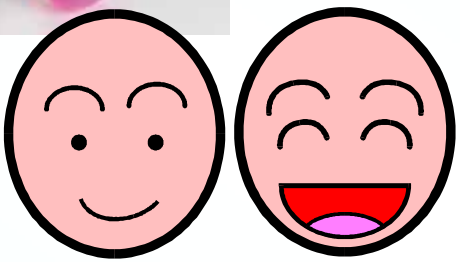
Other Examples

- Other examples are…
  - Stock price estimation
  - Robot motor control
  - Computer vision
  - Spam filter
  - DNA classification





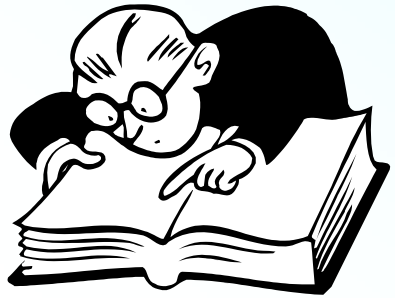




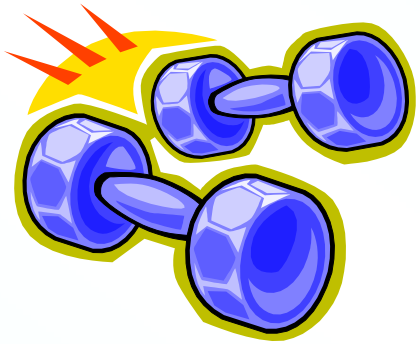
Three Types of Learning

- Supervised learning
  (This course!)

- Unsupervised learning
  (“Advanced data analysis”, 2007 spring)

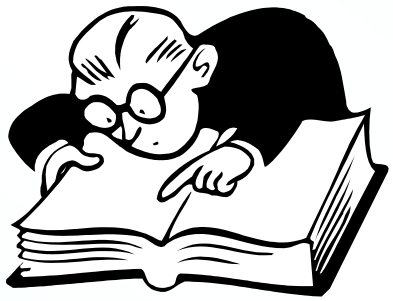


- Reinforcement learning



What Is Unsupervised Learning?

- You are given questions (input data) without answers (output data).
- The goal is to find “interesting” structure in the data.



Interestingness

- The goal of unsupervised learning depends on the definition of “interestingness”:
  - Dimensionality reduction
  - Clustering
  - Blind source separation
  - Outlier detection

Dimensionality Reduction

- Dimensionality reduction (Embedding)
  - We are given high-dimensional data.
  - High-dimensional data is too complex to analyze: Even estimating the density is extremely difficult ("curse of dimensionality")
  - We want to have a low-dimensional expression of the data without losing intrinsic information.
  - Data visualization: Reduced data is less than equal to 3-dimensional.

Example 1

- “Swiss Roll”
- Data is 3-D but it essentially lies on a 2-D manifold.
- We want to “unfold” the roll.

2D

3D

3D

http://www.cs.toronto.edu/~roweis/lle/swissroll.html

Example 2

- Embedding face images into 2D space.
- Images of the same face from different angles and lighting directions (64x64=4096D)

http://isomap.stanford.edu/datasets.html

Example 3

- Embedding hand-written numbers into 2D space.
- Images of different “2” (64x64=4096D)

http://isomap.stanford.edu/datasets.html

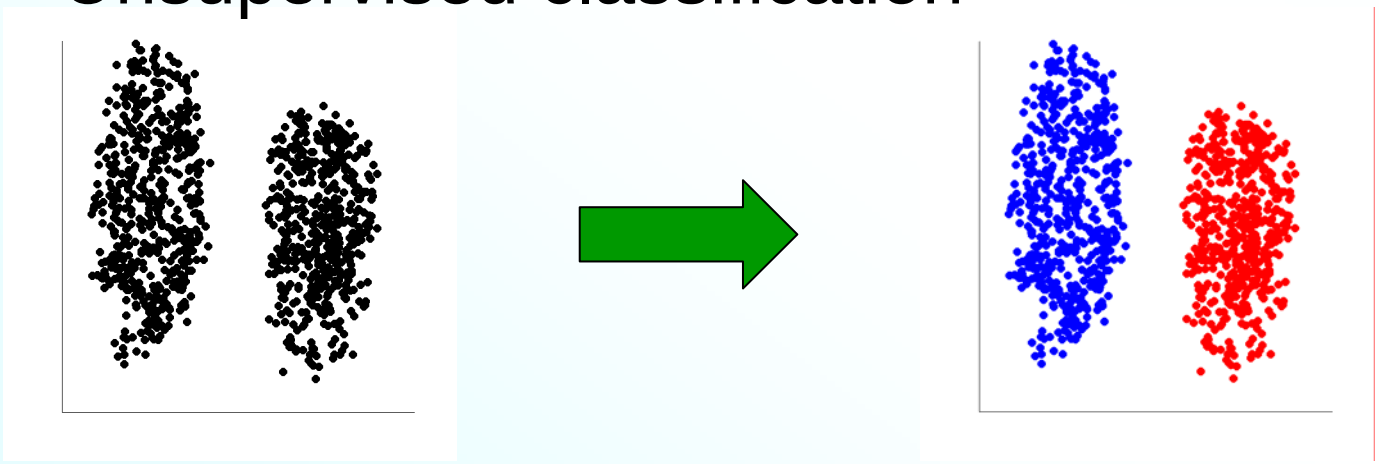
Example 4

- Embedding lip images into 2D space.

http://www.cs.toronto.edu/~roweis/lle/lips.html

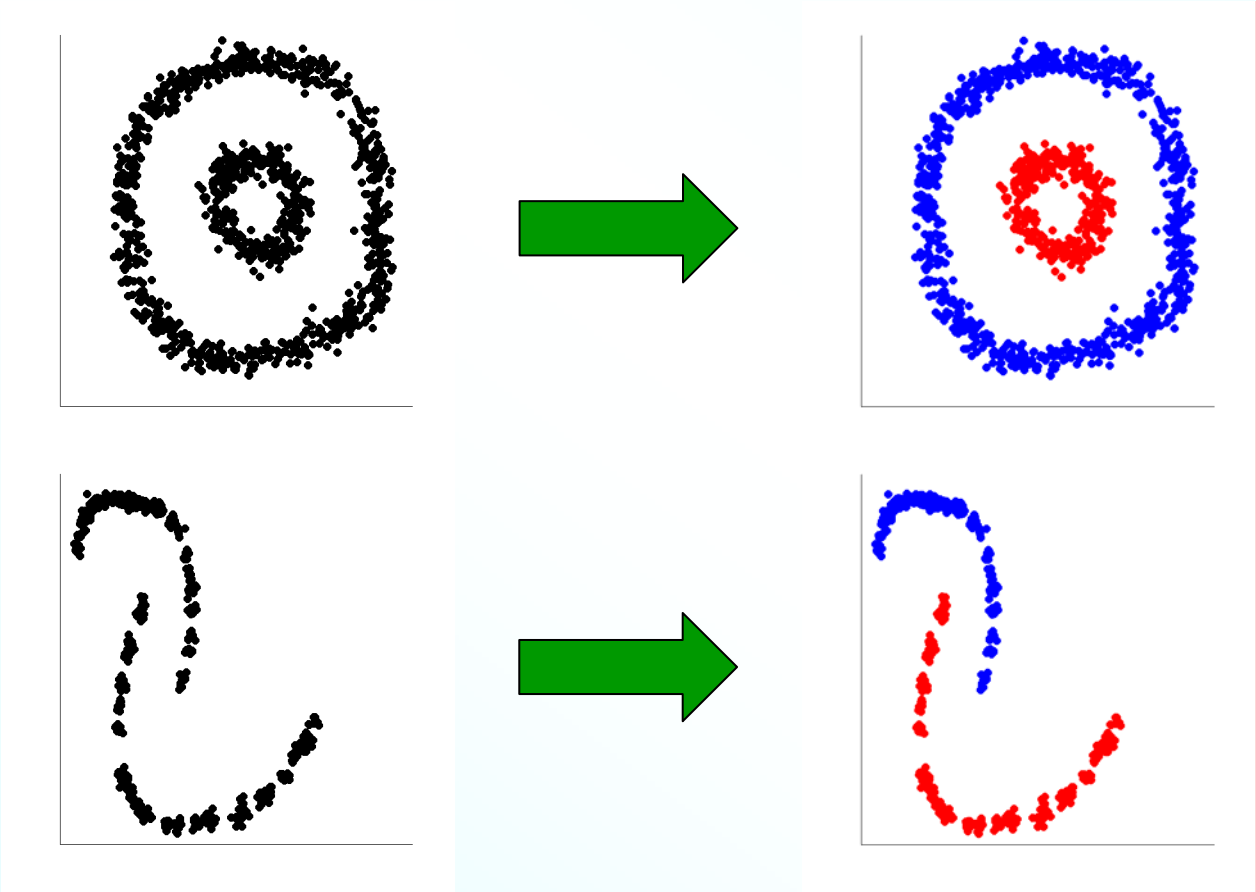
Data Clustering

- Clustering
  - We want to divide the data into disjoint groups so that
    - Data in the same group have similar characteristics.
    - Data in different groups have different characteristics.
  - “Unsupervised classification”



Example 1

- “Connected” points seem to be in the same cluster, rather than “close” points.



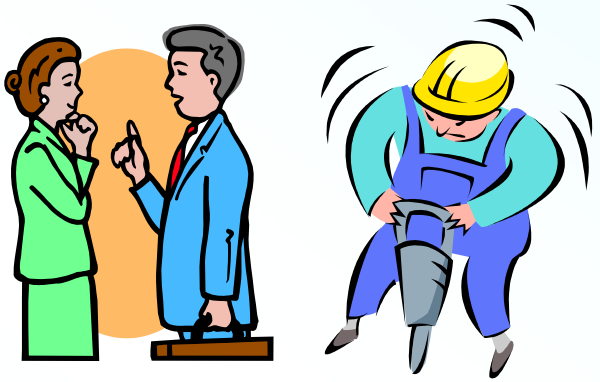
Example 2

- Image segmentation

http://www.cis.upenn.edu/~jshi/software/demo2.html

Blind Source Separation

- We can extract what a person is speaking in a noisy environment.

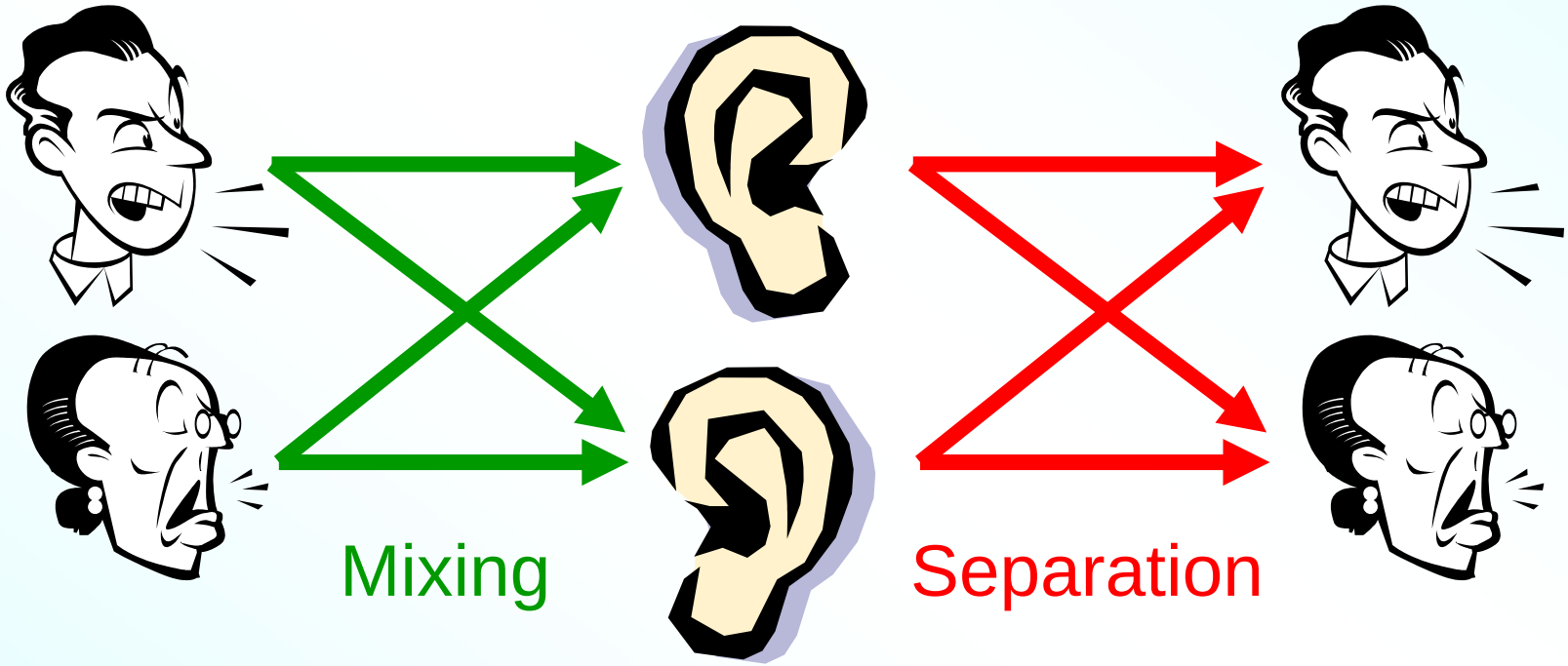


- Syotoku-taishi can distinguish 10 conversations?

From http://www.boj.or.jp/money/now/1man.htm

Blind Source Separation

- Cocktail-party problem

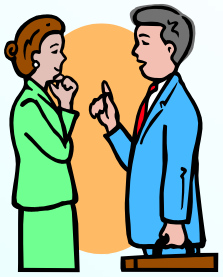


- We want to separate mixed signals into original ones.

Example

	Mixed signal	Separated signal 1	Separated signal 2
Conversation + Conversation			
Conversation + Instrument			

From http://www.brain.kyutech.ac.jp/~shiro/research/blindsep.html



Outlier Detection

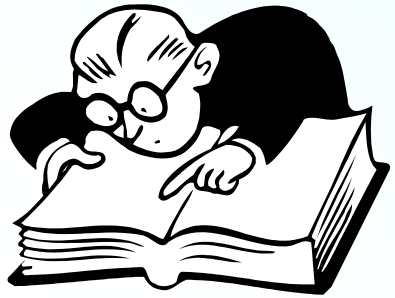
- When a new data sample is added, we want to know whether it is different from the samples collected so far.
- Also referred to as novelty detection, one-class classification

B. Schölkopf et al. (Neural Comp. 2001)

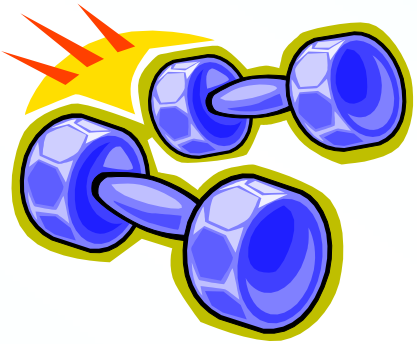
Three Types of Learning

- Supervised learning
  (This course!)

- Unsupervised learning
  (“Advanced data analysis”,
  2007 spring)

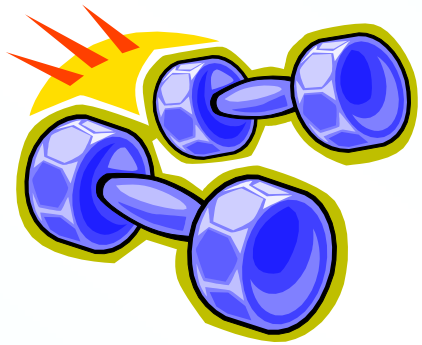


- Reinforcement learning
  (Prof. Shigenobu Kobayashi,
  Dept. of Computational intelligence
  and Systems Science)



What Is Reinforcement Learning?

- The goal of reinforcement learning is same as supervised learning, i.e., to estimate an unknown underlying rule.
- However, different from supervised learning, we are not allowed to ask questions to the teacher.
- Instead, we can get rewards (reinforcement signals) for our estimated answer



What Is Reinforcement Learning?

- Practically, we assume that the rule that maximizes the rewards is the underlying rule.
- Under this assumption, the rule is learned so that the rewards is maximized.
- Reinforcement learning can be regarded as being placed between supervised learning and unsupervised learning.

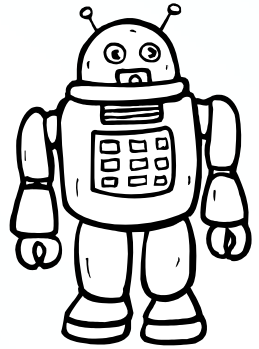
Example

- Learning stand-up motion
- The robot consists of 3 links connected by 2 joints.
- Robot can control it's joint angles by itself.
- The goal is to learn the control rule for stand up.
- Control rule: mapping from inner states to control signal.

From IEICE Trans. Vol. J82-D-II, pp.2118-2131, 1999

Example

- Essentially, reward is given when stand-up motion has been succeeded, otherwise reward is zero.
- However, this does not work well in practice.
- Continuous reward is preferred.
- For example, stand-up is equivalent to lifting the head, the reward is designed such that the higher the head is, the more the reward is.



Example

- Before learning

From http://www.kawato.jst.go.jp/xmorimo/stup_movie.html

Example

- After 750 trials

From http://www.kawato.jst.go.jp/xmorimo/stup_movie.html

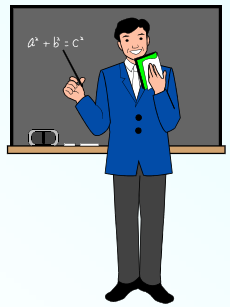
Example

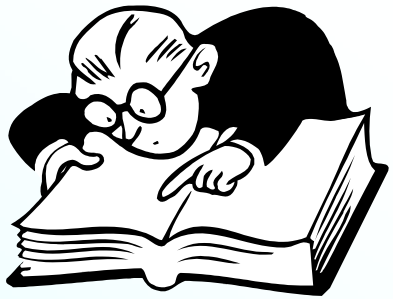
- After 920 trials

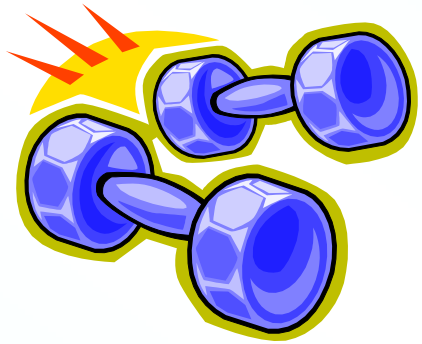
From http://www.kawato.jst.go.jp/xmorimo/stup_movie.html

We Have Learned …

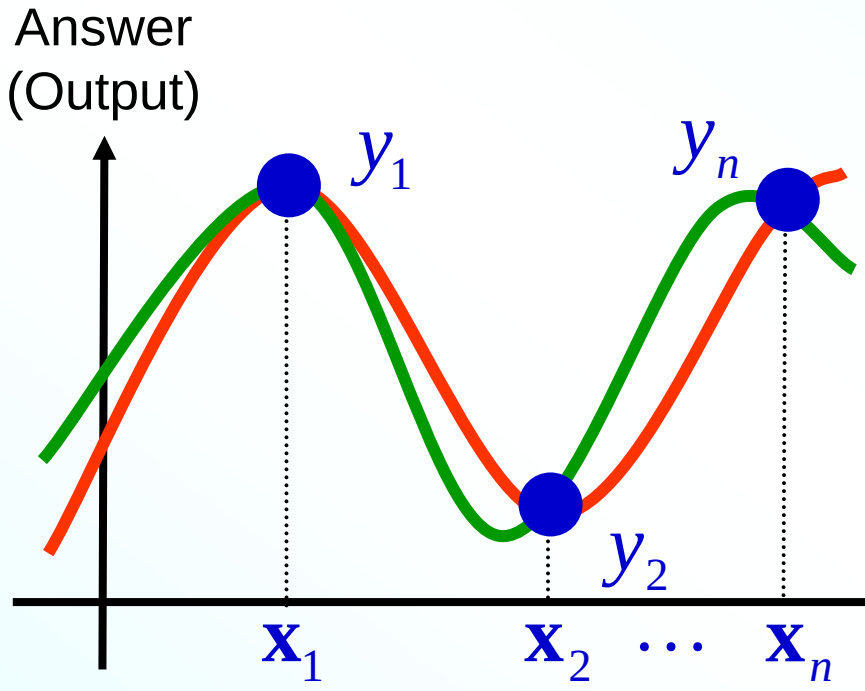
- There are three issues in learning:
  - Supervised learning
  - Unsupervised learning
  - Reinforcement learning







Supervised Learning As Function Approximation



$f(\mathbf{x})$: Target function

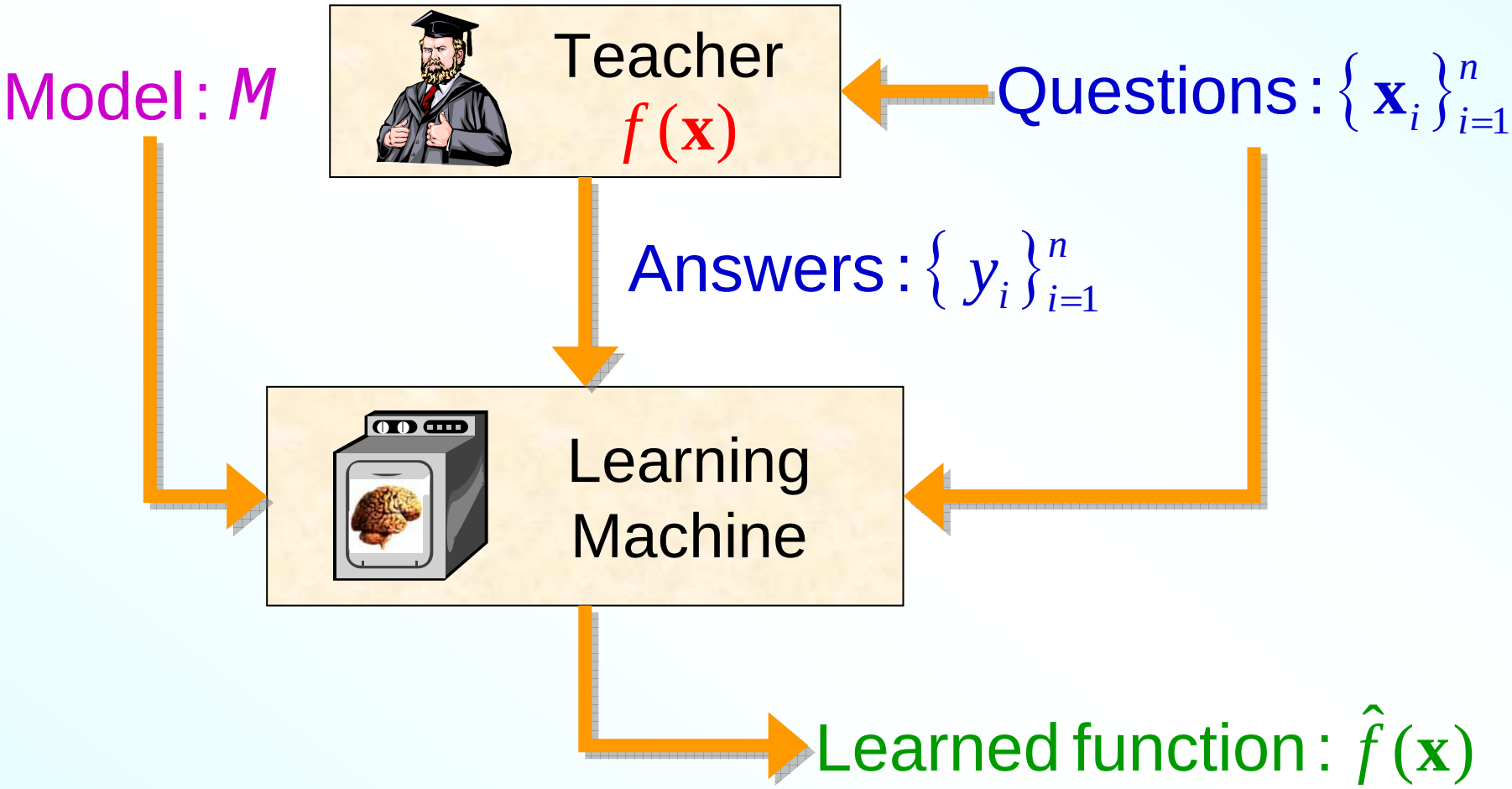
$\hat{f}(\mathbf{x}) \in M$: Learned function

$\{\mathbf{x}_i, y_i\}_{i=1}^{n}$: Training examples

M : Model

Using training examples $\{\mathbf{x}_i, y_i\}_{i=1}^{n}$,
find a function $\hat{f}(\mathbf{x})$ from a model M
that well approximates the target function $f(\mathbf{x})$.

Diagram of Supervised Learning

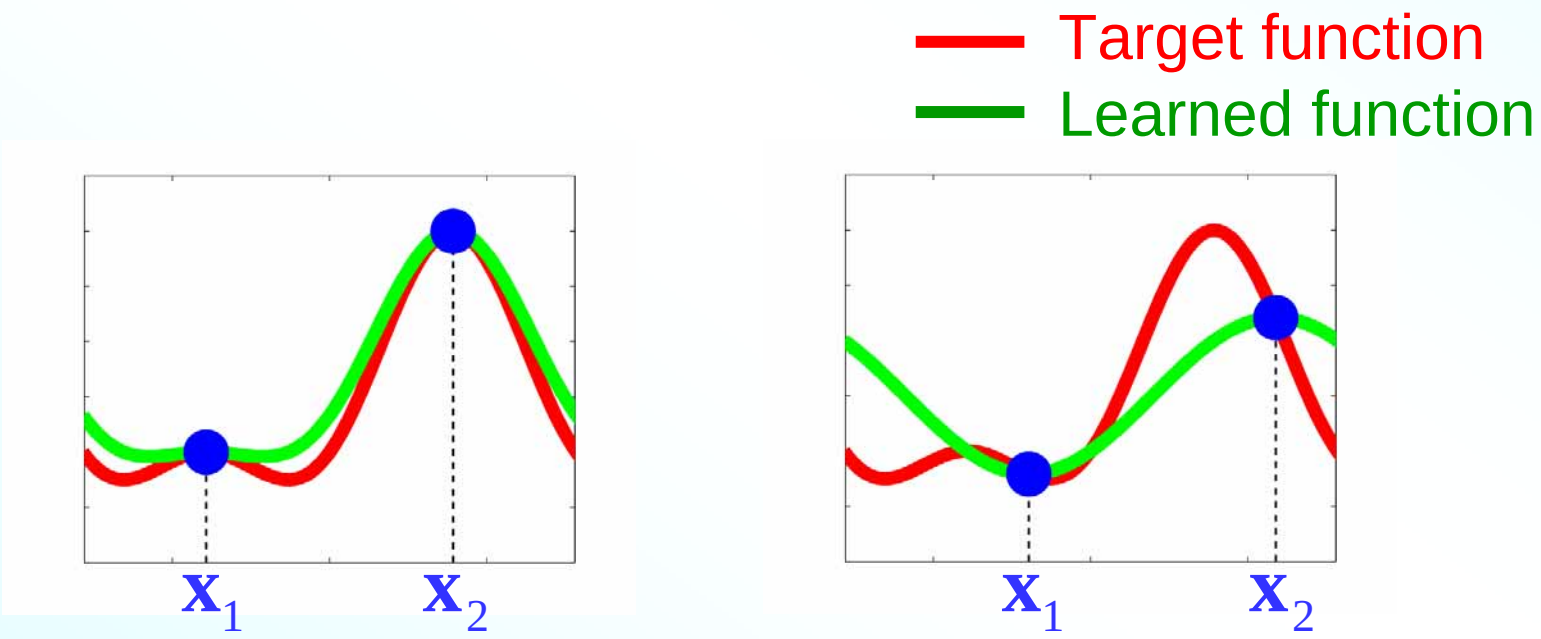


3 Important Topics in Supervised Learning

- Active learning:
  What are the best questions to ask?
- Model selection:
  What is the best model to use?
- Learning methods:
  What is the best way to learn?

Active Learning

For obtaining good learning results,
questions should be determined appropriately.

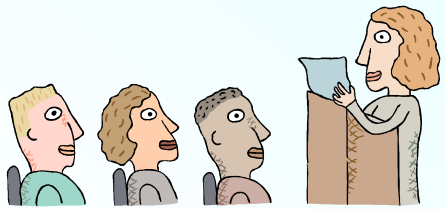


Good questions

Bad questions

Active Learning: Analogy to Real Life

- It is not interesting to passively attend the lecture.



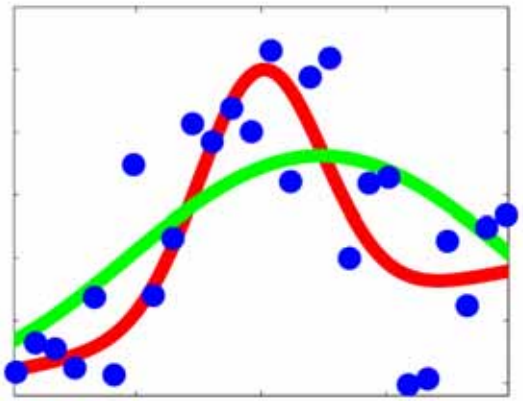
- It is more effective to actively ask questions in the lecture.

Model Selection

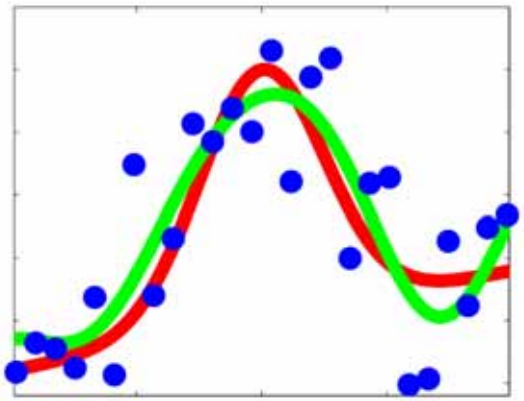
For obtaining good learning results,
model should be determined appropriately.

Model is a set of function
from which learning results are searched.

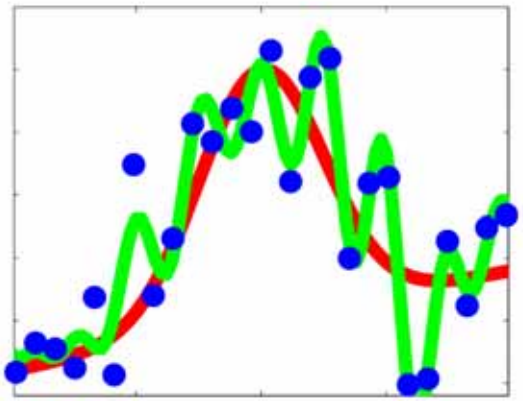
Target function
Learned function



Simple model



Appropriate model



Complex model

Model Selection: Analogy to Real Life

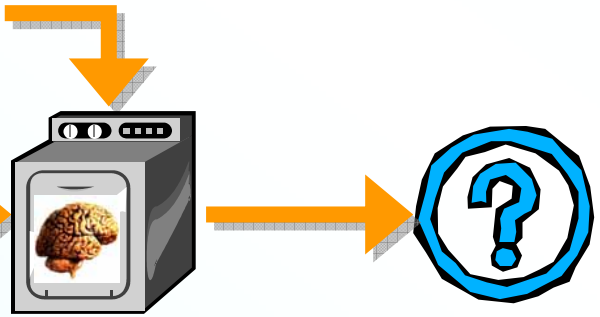
- A model represents your ambition.
- You learn a fixed amount of material.
- If you are less ambitious, you are not capable of even memorizing what you have learned. Therefore, you can not find the truth.
- If you are too ambitious, you can memorize what you have learned perfectly. However, you can not get the whole picture.
- If you are appropriately ambitious, then you can understand the truth.

Learning Methods

- Now you have
  - A model, from which your learning result function is searched.
  - Training examples, which are pairs of questions and their answers.
- A learning method is a rule to specify a function in the model based on the training examples.

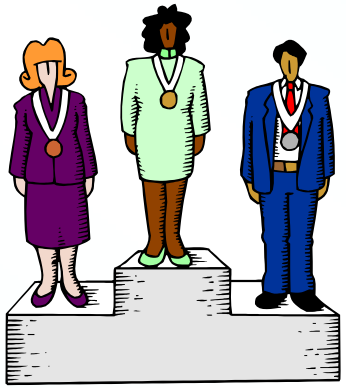
Training examples : $\{\mathbf{x}_i, y_i\}_{i=1}^{n}$

Model : M



Learning Methods: Analogy to Real Life

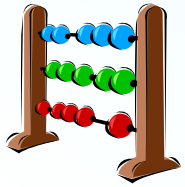
- Now you have
  - Appropriate ambition for learning
  - Good questions and their answers
- What you should do is to just start studying!
- Effectively using your ambition and teaching materials is the key to success.



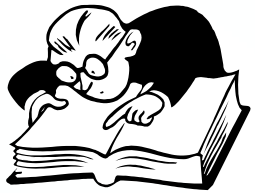
Conclusions

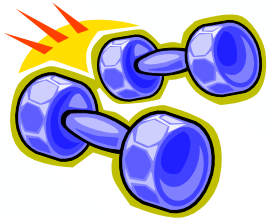
- There are 3 topics in learning research.
  - Understanding human brains
  - Developing learning machines
  - Mathematically clarifying mechanism of learning

- There are 3 types of learning.
  - Supervised learning
  - Unsupervised learning
  - Reinforcement learning



- Topics of supervised learning:
  - Active learning
  - Model selection
  - Learning methods





Homework

- Prepare your own supervised learning data sets (e.g., from your research domain)

$$\{(\boldsymbol{x}_i, y_i)\}_{i=1}^{n}$$

- Input should $\boldsymbol{x}_i$ be vectors and output y_i should be scalars.

$$\boldsymbol{x}_i \in \mathbb{R}^d \qquad y_i \in \mathbb{R}$$

- Better if
  - Input is not so high dimensional
  - Many samples

$$d : \text{small} \qquad n : \text{large}$$

- Explain specification of your data

Homework (cont.)

- Prepare a computer environment in which you can run

  e.g., MATLAB, octave, scilab, R…
- Deadline: beginning of next class (April 25th)