

# Artificial Intelligence — — Introduction



Yanghui Rao

Assistant Prof., Ph.D

School of Data and Computer Science,

Sun Yat-sen University

[raoyangh@mail.sysu.edu.cn](mailto:raoyangh@mail.sysu.edu.cn)

# Instructors

- Course Instructor

- 饶洋辉

- Email: raoyangh@mail.sysu.edu.cn
    - Phone: 15989787592
    - Website: <http://sdcs.sysu.edu.cn/node/2471>

# Announcement of lab course

- Will start in the Third Week
- Five or six experiments
- A final project
- Strictly prohibited for plagiaristic activities  
(严禁抄袭)

# Announcement of lab course

周	课程安排（初步计划）
3	数据集处理
4~5	K近邻算法 朴素贝叶斯算法
6	感知机算法
7~8	决策树
9~10	逻辑回归
11~12	神经网络
13~18	Project

# 实验课要求

- 实验课需要一定的**编程基础**以及**数学基础**，从对公式的推导再到代码的实现，都会在实验课内容中体现。
- **编程基础**包括：
  - 熟练掌握C/C++、JAVA、Python等任意一门编程语言。对语言熟悉的意思是要可以用该门语言的基础编程语句完成所有的编程，并不是调用现成函数包。
  - 人工智能模型强调运算效率。模型的代码实现是基本要求。在此基础上，会要求大家对模型代码进行优化设计，并且进行效率上的比较和分析。

# 实验课要求

- **数学基础包括但不限于：**
  - 熟悉如何对复合函数的特定自变量进行求导，对复杂函数的特定积分元进行积分。
  - 概率学基础知识（如条件概率的计算等），对经典概率函数有所了解（正态分布，高斯分布等）。
  - 线性代数基础知识。包括如何解矩阵方程组，常用矩阵运算的实现，稀疏矩阵的定义等。
  - 了解一些常用的数学知识，如最大似然估计算法原理，余弦距离计算方法，线性回归等各种回归算法的原理等（不作强制要求）。
- **实验课主要包括两项内容：指导实验内容以及验收之前一次的实验内容，有不定时签到，验收会包括推导公式，解释代码以及现场跑结果。**

# Background

- **Camera:** Simulate part functions of eyes
  - images and videos -> electrical signal
  - Computer Graphics (CG), Computer Vision (CV), etc.
- **Microphone:** Simulate part functions of ear
  - audios -> electrical signal
  - Speech Recognition (SR), etc.
- **Computer:** Simulate small part functions of brain
  - computing brain; symbol -> pattern
  - Artificial Intelligence (AI), etc.
- ...
- The connection between digital devices and brain
  - Human-Computer Interaction (HCI)



# Human-Computer Interaction

- Substitute some functions of human, make *human* better





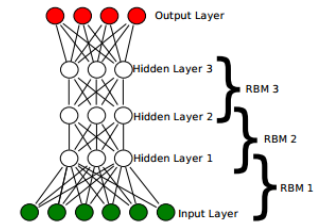
# Artificial Intelligence

- Simulate more functions of brain, make *computer* better
  - chess brain, question answering brain, drive brain, etc.

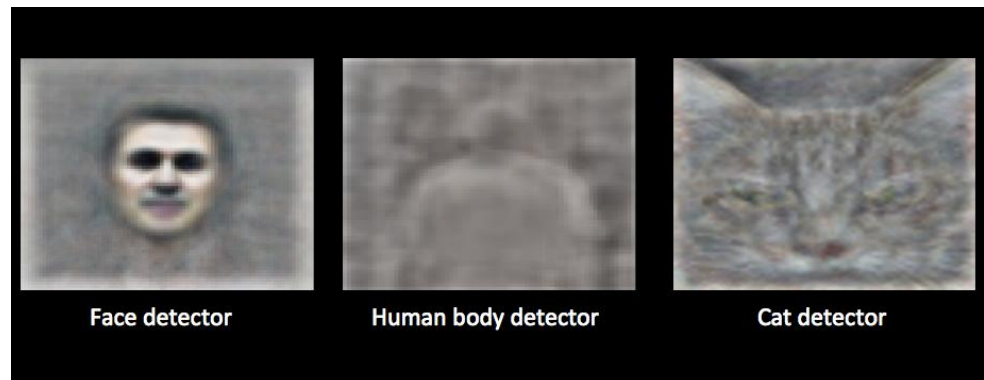


# Artificial Intelligence

- Simulate more functions of brain, make *computer* better
  - chess brain, question answering brain, drive brain, etc.



- face detector
- human body detector
- cat detector
- ...



# History of AI

- Turing Test, proposed by A.M. Turing in 1950
- 1950年，阿兰·图灵提出图灵测试，为智能提供一个满足可操作要求的定义。图灵测试用人类的表现来衡量假设的智能机器的表现，这无疑是评价智能行为的最好且唯一的标准。



# History of AI

- 图灵称为“模仿游戏”的测试是这样进行的：将一个人与一台机器置于一间房间中，而与另外一个人分隔开来，并把后一个人称为询问者。询问者不能直接见到屋中任一方，也不能与他们说话，因此，他不知道到底哪一个实体是机器，只可以通过一个类似终端的文本设备与他们联系。
- 然后，让询问者仅根据通过这个仪器提问收到的答案辨别出哪个是计算机，哪个是人。如果询问者不能区别出机器和人，那么根据图灵的理论，就可以认为这个机器是智能的。

# History of AI

- Turing Test, proposed by A.M. Turing in 1950
- Loebner Prize, pledged by Hugh Loebner in 1990
  - “Question Answering”/Chat robots
  - <http://www.loebner.net/Prizetf/loebner-prize.html>
  - 1991, Joseph Weintraub
    - If you say "I need my Mommy", ELIZA will say "Tell me why you need your Mommy?"
    - It employs AI sentence parsing and knowledgebase technology, plus a 70,000 word vocabulary.
  - ...

# History of AI

- Logic, 1950s
  - “Knowledge” is transferred by experts/humans
  - *propositional* logic (命题逻辑), first-order *predicate* logic (谓词逻辑), *fuzzy* logic...



# Logic

- “true” or “false”
- negation
- conjunction
- disjunction
- implication
- equivalence
- universal quantifier
- existential quantifier
- ...

```
~~~~~欢迎您!~~~~~
      游戏：五子棋
~~~~~

人人对弈1  人机对弈2
请您选择1  或者2:  2
您选择了人机对弈~~~~
黑方1  US  白方2
请您选择1  或者2:  2
您选择了2,对方先下~~

  1 2 3 4 5 6 7 8 9 A B C D E F
1 + + + + + + + + + + + + + +
2 + + + + + + + + + + + + + +
3 + + + + + + + + + + + + + +
4 + + + + + + + + + + + + + +
5 + + + + + + + + + + + + + +
6 + + + + + + + + + + + + + +
7 + + + + + + + + + + + + + +
8 + + + + + + + + + + + + + +
9 + + + + + + + + + + + + + +
A + + + + + + + + + + + + + +
B + + + + + + + + + + + + + +
C + + + + + + + + + + + + + +
D + + + + + + + + + + + + + +
E + + + + + + + + + + + + + +
F + + + + + + + + + + + + + +
该计算机下了, 请输入1
```

# Logic

- “true” or “false”
- negation
- conjunction
- disjunction
- implication
- equivalence
- universal quantifier
- existential quantifier
- ...

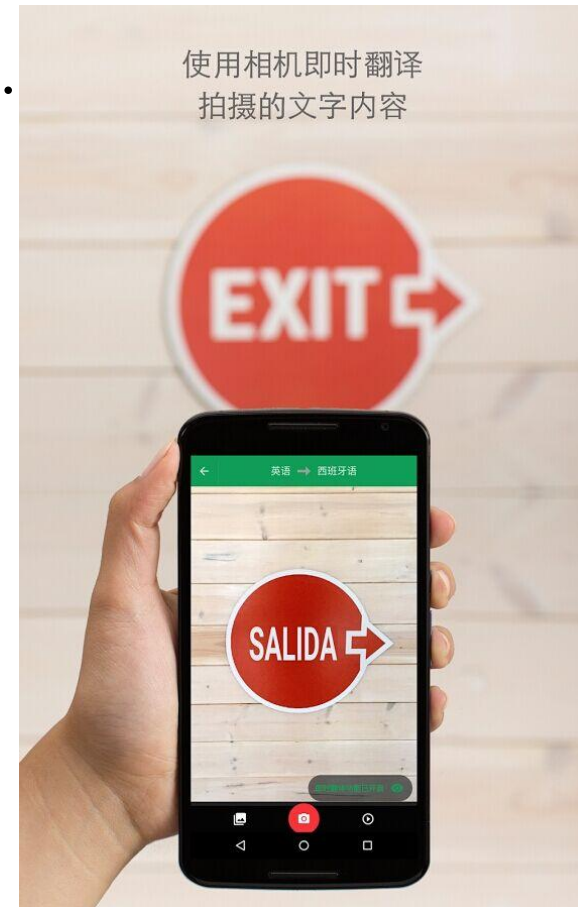
```
当前步数为: 40
  1 2 3 4 5 6 7 8 9 A B C D E F
1 + + + + + + + + + + + + + +
2 + + + + + + 0 0 0 + + + + +
3 + + 0 0 0 0 0 0 0 + + + + +
4 + + + 0 0 0 0 0 0 + + + + +
5 + + + 0 0 0 0 0 0 + + + + +
6 + + 0 0 0 0 0 0 0 + + + + +
7 + 0 0 0 0 0 0 0 + + + + +
8 + + + 0 + 0 0 + + + + +
9 + + + 0 + + + + + + + + +
A + + + + + + + + + + + + +
B + + + + + + + + + + + + +
C + + + + + + + + + + + + +
D + + + + + + + + + + + + +
E + + + + + + + + + + + + +
F + + + + + + + + + + + + +
您的输入是: 6 3
祝贺您!您赢了~
```

# History of AI

- Logic, 1950s
  - “Knowledge” is transferred by experts/humans
  - *propositional* logic (命题逻辑), first-order *predicate* logic (谓词逻辑), *fuzzy* logic...
- Machine Learning, 1990~
  - “Knowledge” is learnt by computers primarily
  - *supervised* learning (监督学习): 分类, 回归...
  - *unsupervised* learning (无监督学习): 关联规则挖掘, 聚类...
  - *reinforcement* learning (强化/增强学习)

# What is Artificial Intelligence

- Compute something that shows **intelligent** behavior
  - Natural language processing, ...



# What is Artificial Intelligence

- Compute something that shows **intelligent** behavior
  - Natural language processing, ...

机器人Pepper和阿里巴巴集团零售事业群总裁张建锋互动：

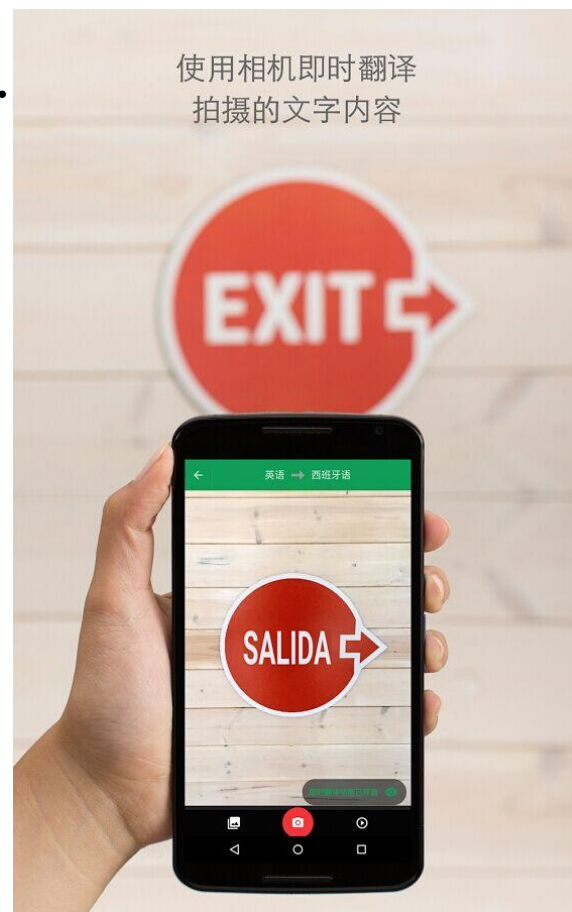
.....

Pepper: 快乐20, 紧张60, 生气0, 悲伤0, 我的情感引擎显示, 您好象有点紧张, 是吗?

.....

Pepper: 快乐80, 紧张20, 生气0, 悲伤0。我的情感引擎显示你的心情似乎不错。你是不是开始喜欢我了?

.....



# 人工智能

- 1956年夏季，人工智能（AI）作为一门独立的学科正式诞生在达特茅斯大学召开的世界上第一次人工智能大会。经McCarthy提议，在会上正式决定使用人工智能一词来概括这个研究方向。
- MIT、斯坦福大学和CMU被称为人工智能和计算机科学的三大中心。
- 人工智能研究的主要内容包括机器学习、模式识别、自然语言处理、智能决策支持系统、问题求解、专家系统等广泛内容。

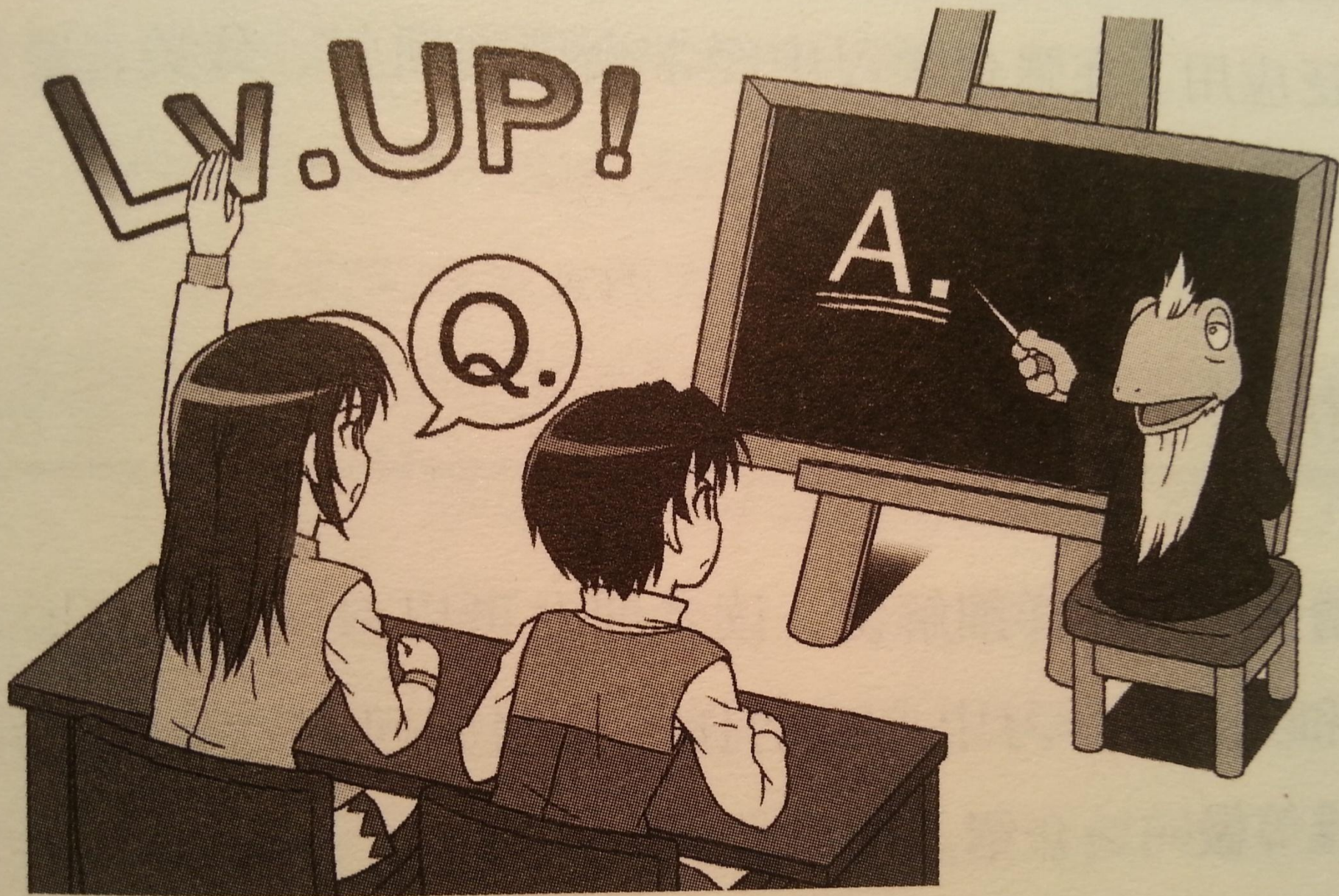


# What is Machine Learning

- Tom Mitchell (1997): A **computer program** (vs *human*) is said to **learn** from experience  $E$  (e.g., **labeled data**) with respect to some class of tasks  $T$  (e.g., **classification**) and performance measure  $P$  (e.g., **precision & recall**), if its performance at tasks in  $T$ , as measured by  $P$ , improves with experience  $E$ .

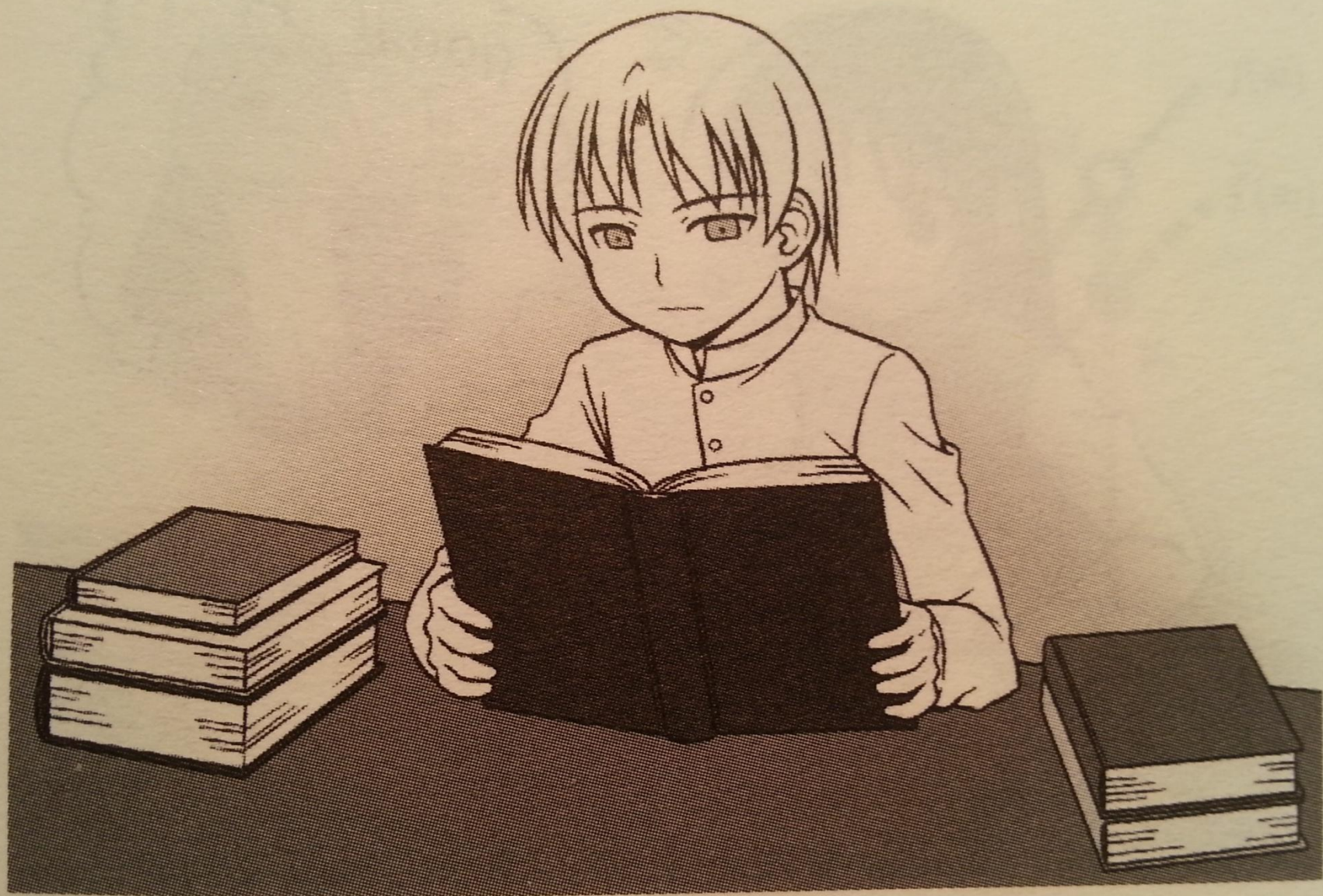
# 机器学习

- 机器学习就是让计算机能够像人那样自动获取新知识，并在实践中不断地完善自我和增强能力，使得系统在下一次执行同样任务或类似的任务时，会比现在做得更好或效率更高。
- 机器学习的研究一方面可以使机器能自动获取知识，赋予机器更多的智能；另一方面可以进一步揭示人类思维规律和学习奥秘，帮助人们提高学习效率。机器学习的研究还会对记忆存储模式、信息输入方式及计算机体系结构产生重大影响。



监督学习





无监督学习





强化学习

# What is Pattern Recognition

- Christopher M. Bishop (2006): The automatic discovery of regularities in data through the use of computer algorithms (*e.g.*, Machine Learning) and with the use of these regularities to take actions such as classifying the data into different categories



# 模式识别

- 模式识别是研究如何使机器具有感知能力，主要研究视觉模式和听觉模式的识别。如识别物体、地形、图像、字体（如签字）等。
- 模式识别是一个不断发展的新学科。它的理论基础和研究范围也在不断发展。随着生物医学对人类大脑的初步认识，模拟人脑构造的计算机实验即人工神经网络方法早在20世纪50年代末和60年代初就已经开始。在日常生活各方面以及军事上都有广大的用途。至今，在模式识别领域，神经网络方法已经成功地用于手写字符的识别、汽车牌照的识别、指纹识别、语音识别等方面。

# What Kinds of Regularities?

- **Classification**

- A loan user: high or low risk?  $\leftrightarrow$  *banker*
- A person: health or sick?  $\leftrightarrow$  *doctor*
- An Iris flower: Setosa, Versicolour, or Virginica?  $\leftrightarrow$  *botanist* (植物学家)



# Classification

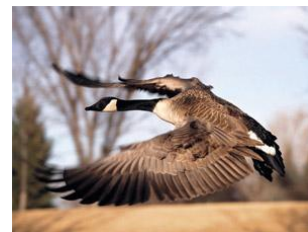
- Make the computer as intelligent as an expert
  - Classify a person as healthy or sick
  - Identify the author of a piece of art or a book
  - Identify the variety of an animal or a plant

# Classification

- Make the computer as intelligent as an expert
  - Classify a person as healthy or sick
  - Identify the author of a piece of art or a book
  - Identify the variety of an animal or a plant
- Training data: examples of the input vectors along with their corresponding target vectors
  - Input vectors: height, weight, has a tail? (yes or no)
  - Target vectors: human or monkey

# Classification

- Make the computer as intelligent as an expert
  - Classify a person as healthy or sick
  - Identify the author of a piece of art or a book
  - Identify the variety of an animal or a plant
- Training data: examples of the input vectors along with their corresponding target vectors
  - Input vectors: height, weight, has a tail? (yes or no)
  - Target vectors: human or monkey
  - Which attributes are good to classify a bird from others?



# Classification

- Input vectors of birds:
  - can move? (yes or no)
  - can chirp? (yes or no)
  - have feather? (yes or no)
  - size (length, width, height)
  - reaction to sound? (yes or no)
  - ...



# Classification

- Input vectors of birds:
  - can move? (yes or no)
  - can chirp? (yes or no)
  - have feather? (yes or no)
  - size (length, width, height)
  - reaction to sound? (yes or no)
  - ...



# Classification

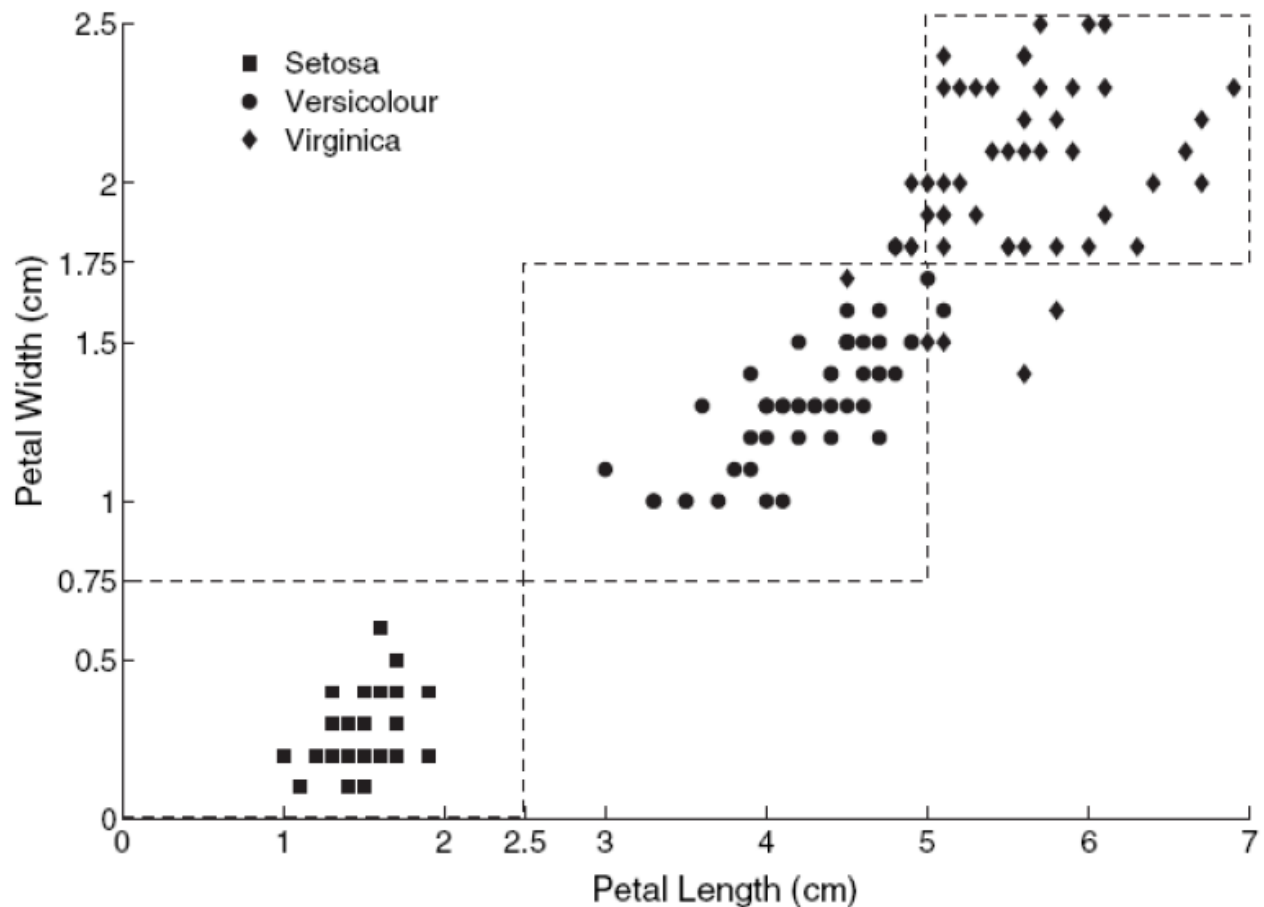
- Data set for Machine Learning
  - UCI Machine Learning Repository (<http://archive.ics.uci.edu/ml/datasets.html>)
- Iris data set
  - <http://archive.ics.uci.edu/ml/datasets/Iris>
  - Attributes provided by Fisher (experts)
    - sepal (萼) length (cm), sepal width (cm), petal (瓣) length (cm), petal width (cm), class

5.3,3.7,1.5,0.2,	Iris-setosa
5.0,3.3,1.4,0.2,	Iris-setosa
7.0,3.2,4.7,1.4,	Iris-versicolor
6.4,3.2,4.5,1.5,	Iris-versicolor
6.3,3.3,6.0,2.5,	Iris-virginica
5.8,2.7,5.1,1.9,	Iris-virginica

**Classifier:** Decision tree, SVM, Neural Network, ...

# Classification

- Plot using petal length and width



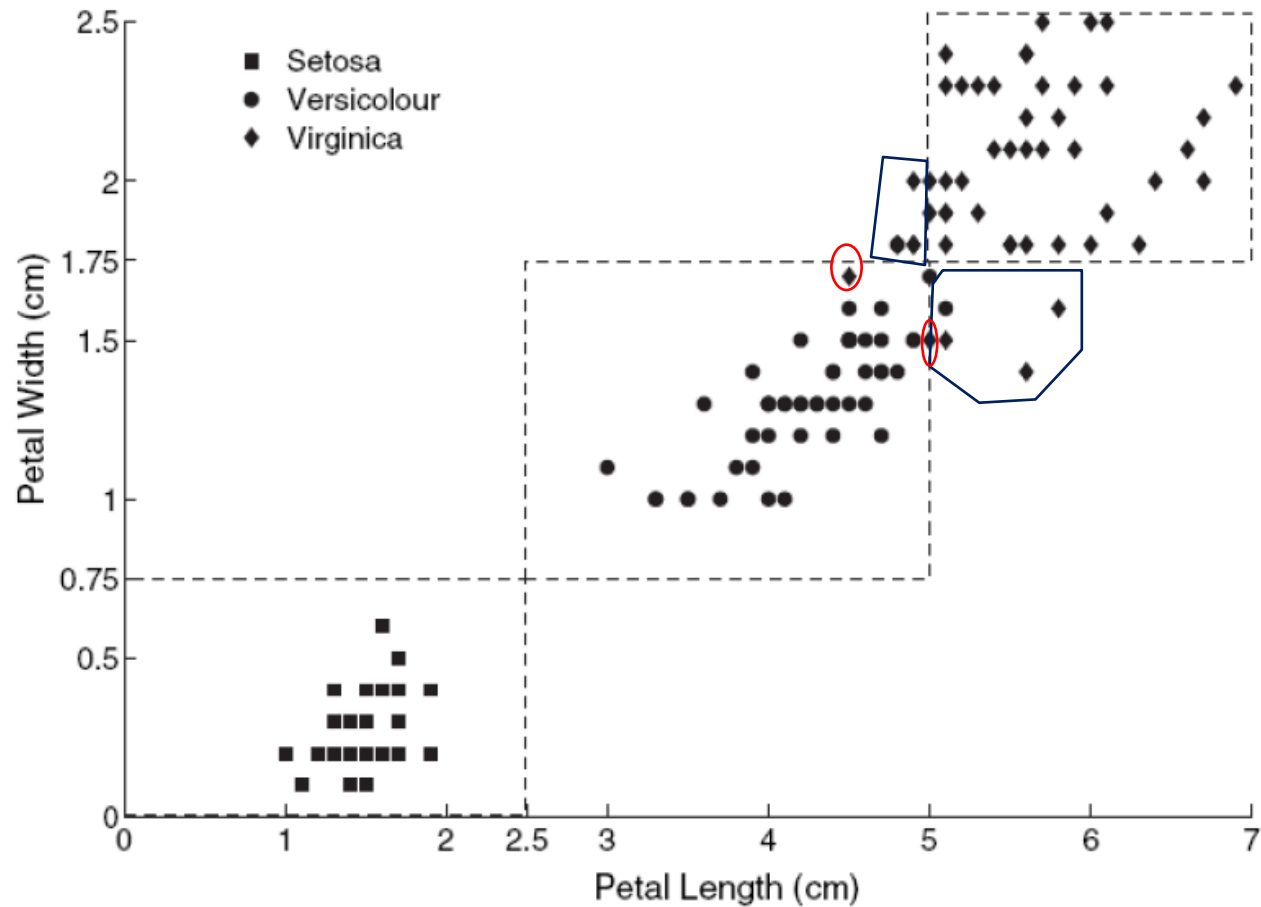
# Classification

- Based on these categories with data, the following rules can be derived:
  - Petal width low and petal length low implies Setosa (山鸢尾).
  - Petal width medium and petal length medium implies Versicolour (变色鸢尾).
  - Petal width high and petal length high implies Virginica (维珍尼亚鸢尾).

# Classification

- These rules do not classify all the flowers correctly.
  - Flowers from the Setosa species are well separated from the other two species with respect to petal width and length.
  - However, the Versicolour and Virginica species overlap somewhat with respect to these two attributes.
- Solutions?

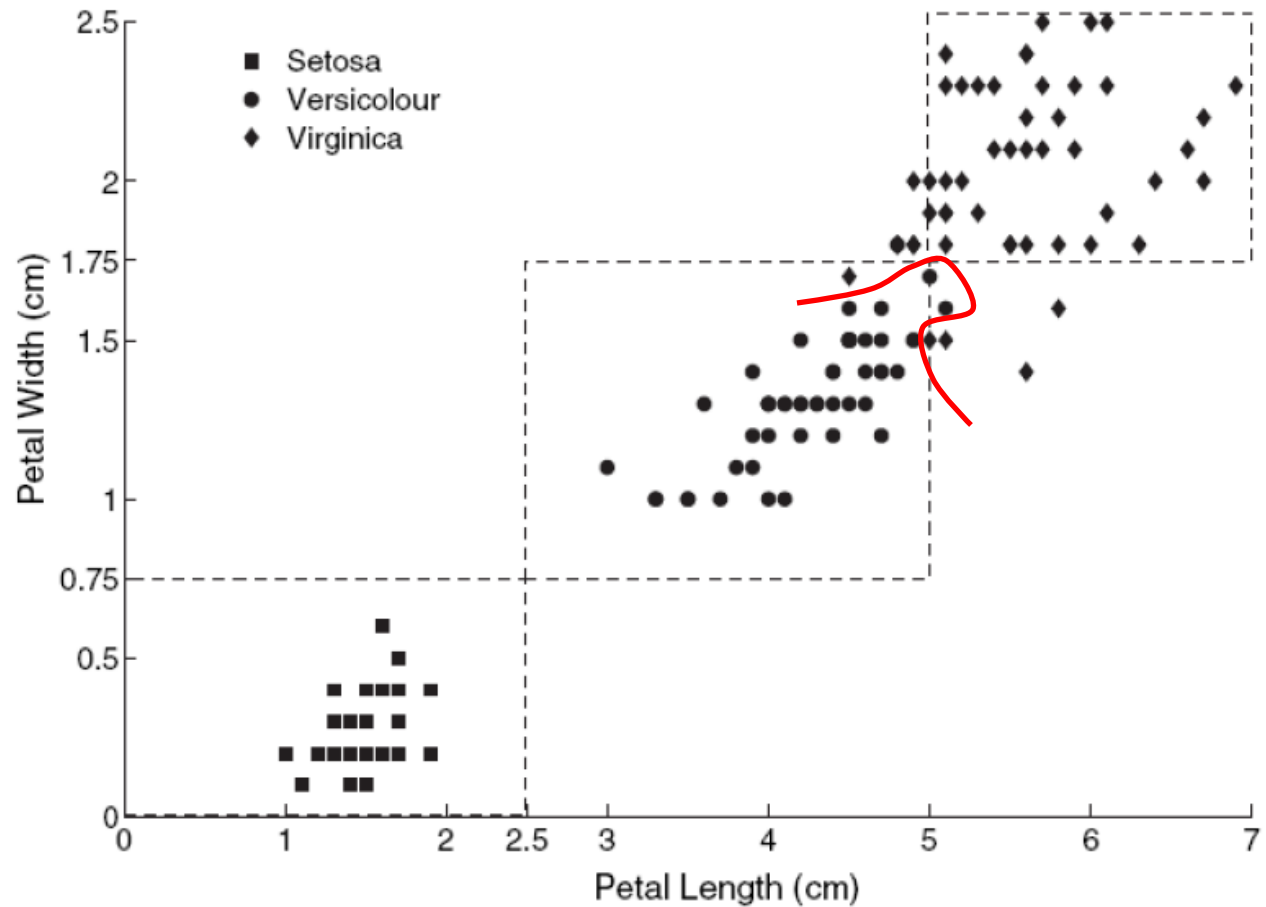
# Classification



1. Records surrounded by red line: misclassified
2. Records surrounded by blue line: unclassified

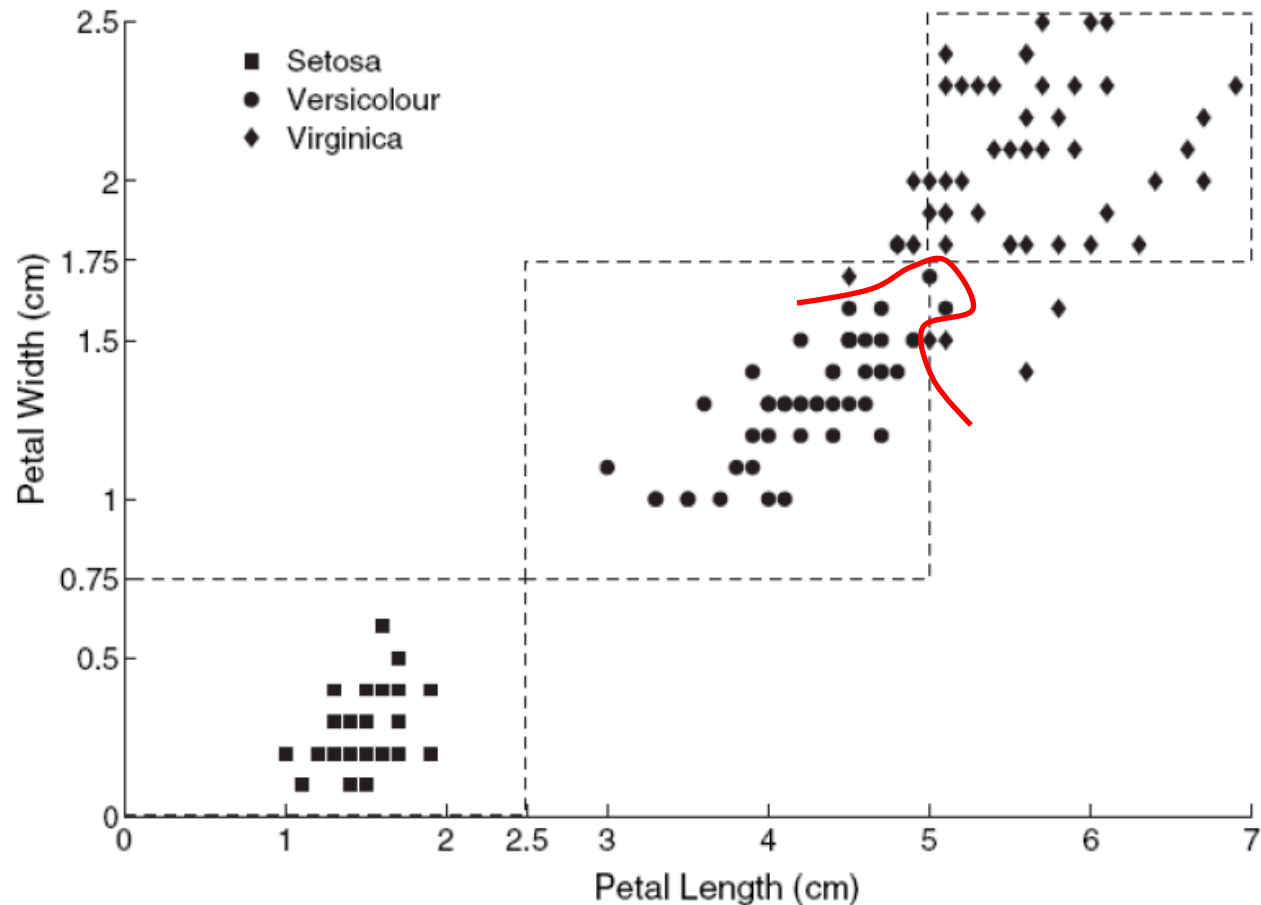


# Classification



1. A new better classifier model

# Classification



1. A new better classifier model
2. Add more attributes  
(i.e., sepal length and width, and others if possible)

# What Kinds of Regularities?

- **Regression**

- Highly related to statistics
- Many concerned quantitative variables:
  - A person's life expectancy
  - A journal article's citations
  - A city's air temperature next day
  - ...
- Mining the correlation between a series of independent variables ( $X$ ) and dependent variables ( $Y$ )

# Regression

- Which one of the following is false for building a regression model?

(X: independent variables; Y: dependent variables)

- 1. X is the working years of an employee, Y is the employee's salary
- 2. X is a child's height, Y is the height of the child's father or mother
- 3. X is the total production of a product, Y is the total consumption of the product
- 4. X is the total consumption of a product, Y is the total production of the product

# Regression

- Which one of the following is false for building a regression model?

(X: independent variables; Y: dependent variables)

- 1. X is the working years of an employee, Y is the employee's salary
- 2. X is a child's height, Y is the height of the child's father or mother
- 3. X is the total production of a product, Y is the total consumption of the product (**iPhone**)
- 4. X is the total consumption of a product, Y is the total production of the product

# What Kinds of Regularities?

- Association Rules
  - Detect sets of attributes or items that frequently co-occur in many records
  - On Thursday, during 4-11pm, customers often purchase diapers and beers together





# Association Rules

- Where does the data come from?
  - supermarket transactions, customer complaint calls, discount coupons, etc.
- Cross-marketing analysis
  - purchase recommendation, cross selling
  - what are the subsequent purchases after buying a given product?
- Target-marketing
  - what types of users buy what products
- Catalog design

# What Kinds of Regularities?

- **Clustering**

- seeks to find groups of closely related records (*e.g.*, users, articles, genes, planets)
- can be applied to
  - compress data
  - anomaly detection
  - reduce dimensions
  - group sets of related customers/products

# Clustering

- Key steps
  - Feature extraction for each record
    - A user of Amazon
    - An article, *etc.*
  - Weighting for each feature
    - Linear
    - Nonlinear, *etc.*
  - Similarity measurement for *paired* records
    - Euclidean distance
    - Cosine similarity, *etc.*

# Clustering

- How to decide these three articles as two groups/clusters?
  - 1: *I'm involved in the release of apple's iPhone 6.*
  - 2: *Apple's iPhone 6 released on Friday.*
  - 3: *I've eaten an apple this Friday.*

The screenshot shows the Clusty search engine interface. At the top, there's a navigation bar with links for 'web', 'news', 'blogs', 'wikipedia', 'jobs', and 'more'. The search bar contains 'iPhone 6' and a 'Search' button. To the right of the search bar are links for 'advanced preferences'. Below the navigation bar, there's a horizontal menu with categories: 'Top News', 'World', 'U.S.', 'Business', 'Sports', 'Health', 'Tech', and 'Science'. The main content area displays 'Top 307 results retrieved for the query iPhone 6 (details)'. On the left side, there's a sidebar with a 'remix' button and a list of 'All Results' (307) categorized by sources: 'Pre-orders' (28), 'Apple Pay' (21), 'Samsung' (17), 'China' (20), 'Stocks' (17), 'iPhone 6 launch' (15), 'IOS 8' (13), and 'Buy The Iphone 6' (10). The main content area shows two article snippets. The first article is titled 'Apple's latest iPhones: What's in the box?' and is dated '1 hour ago'. It mentions that the new iPhones are bigger, slimmer, and have faster processors, and asks what's in the new iPhone 6 and iPhone 6 Plus box. The second article is titled 'Big Day for Apple's Plus-Sized Phones' and is dated '2 hours ago'. It mentions that September 19 has been circled on the calendar of Apple enthusiasts, who woke up in the middle of the night last week to pre-order the new iPhone 6 and iPhone 6 Plus. Both articles include source links and icons for social media sharing.

Clusty

web news blogs wikipedia jobs more »

iPhone 6 Search advanced preferences

clouds sources sections time remix

All Results (307)

- Pre-orders (28)
- Apple Pay (21)
- Samsung (17)
- China (20)
- Stocks (17)
- iPhone 6 launch (15)
- IOS 8 (13)
- Buy The Iphone 6 (10)

Top 307 results retrieved for the query iPhone 6 (details)

Search Results

Apple's latest iPhones: What's in the box? 1 hour ago - They are bigger, slimmer and have faster processors than the iPhones that came before them, but what's in Apple's new iPhone 6 and iPhone 6 Plus box? [www.cnbc.com/id/102009946](http://www.cnbc.com/id/102009946) - [cache] - CNBC

Big Day for Apple's Plus-Sized Phones 2 hours ago - September 19 has been circled on the calendar of Apple enthusiasts who woke up in the middle of the night last week to pre-order the new iPhone 6 and iPhone 6 Plus. Others have spent days camped outside Apple stores for a chance to be one of the first customers to purchase a new device when the... [abcnews.go.com/Technology/buying-apple-iphone-iphone-today/story?id=25601151](http://abcnews.go.com/Technology/buying-apple-iphone-iphone-today/story?id=25601151) - [cache] - ABC News

# 自然语言处理

- 如果能让计算机“听懂”、“看懂”人类自身的语言(如汉语、英语、法语等),那将使更多的人可以使用计算机,大大提高计算机的利用率。自然语言理解就是研究如何让计算机理解人类自然语言的一个研究领域。自然语言理解是指机器能够执行人类所期望的以下语言功能:
  - 回答有关提问(question-answering)
  - 摘要生成和文本释义(Summarizing and paraphrase)
  - 自动翻译(translation)



# Summary

- 随着计算机技术的快速发展，人工智能的研究近几年来也取得了很多新的进展，许多研究领域不断出现，例如数据挖掘、网络信息过滤等新型的研究领域。随着科学技术的发展，人工智能各研究领域间的联系将更加紧密，互相渗透，这种融合与渗透必将促进人工智能研究的发展，促使其走向实际应用。