Course ID 80250993 Dates: 9/16-12/23/2021
@ I-205 Lotus RainClassroom + Tencent/VooV Meeting

清華大學

Machine Learning

For online students:

Tencent Meeting /Voov Meeting
https://meeting.tencent.com/dm/zTZbFHoW618m?rs=27
ID: 521 4678 6257
Password: 1205
Broadcast (save this link for replay within 30 days):
https://meeting.tencent.com/l/p11V4drTYoa4
Password: 1205

Lotus RainClassroom

二维码有效期截至 2021-09-25

邀请码:RIDFHW

荷塘雨课堂 公众号中输入，长期有效

投票
最多可选1项
设置
提交

Survey in the Rain Classroom

- Ⓐ I’m in the classroom
- Ⓑ I’m in Mainland China but not in the classroom
- Ⓒ I’m in Asia-Pacific but not in Mainland China
- Ⓓ I’m in the Middle-East
- Ⓔ I’m in Europe or Africa
- Ⓕ I’m in America

Test the Tencent Meeting or VooV meeting (https://voovmeeting.com/)

- For those who are online, please turn on your video camera and microphone, and say hello to everyone

---

Course ID 80250993

Machine Learning

Instructor: Xuegong Zhang （张学工）

zhangxg@tsinghua.edu.cn

TAs:

- **Qiuchen Meng 孟秋辰 mqc17@mails.tsinghua.edu.cn**
- Jiaqi Li 李嘉骐 li-jq18@mails.tsinghua.edu.cn
- Xi Xi 席曦 xix19@mails.tsinghua.edu.cn
- Yixin Chen 陈奕鑫 chenyx19@mails.tsinghua.edu.cn
- Minsheng Hao 郝敏升 hms20@mails.tsinghua.edu.cn

TU 80250993 Machine Learning (Fall 2021 Beijing Class)

Chapter 1
Introduction

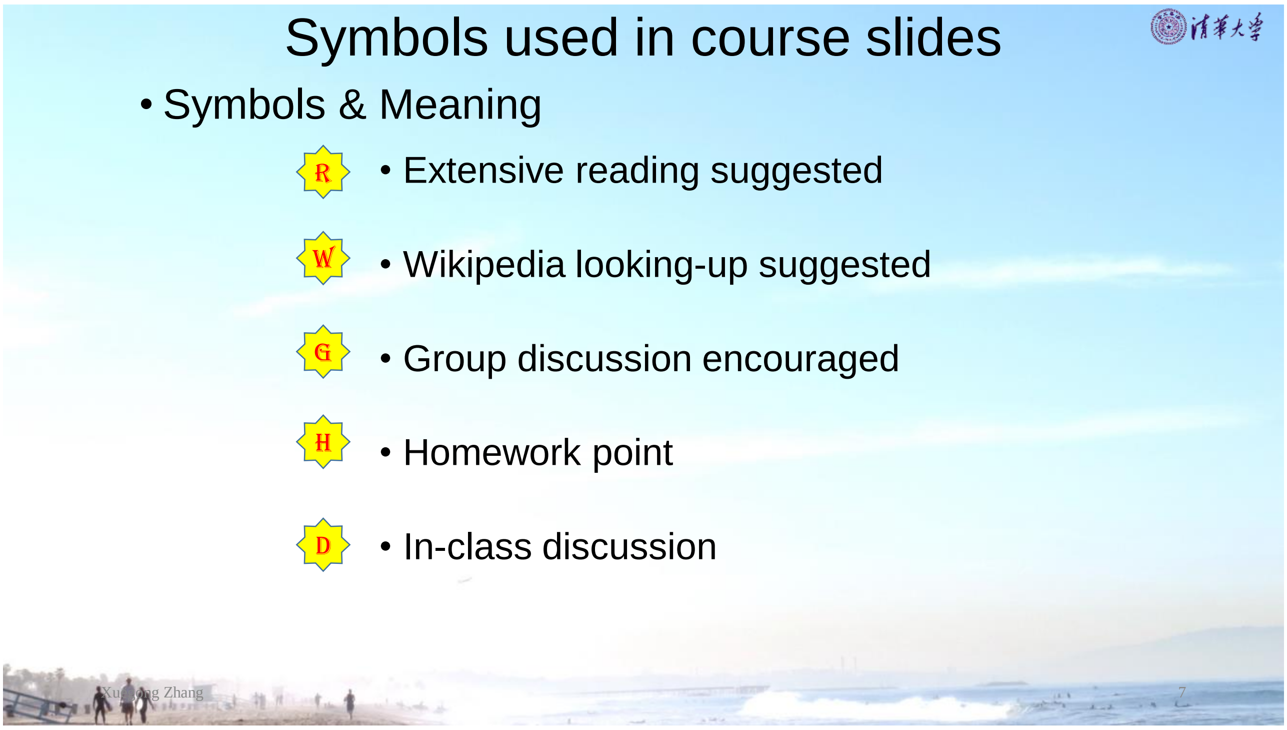
Xuegong Zhang
September 16, 2021

Self Introduction

- Xuegong Zhang
  - Department of Automation, Tsinghua Univ: B.S., 1989; Ph.D., 1994
  - Professor, Department of Automation, TU
  - Adjunct Professor, School of Life Sciences and School of Medicine, TU
  - Visiting Scholar, Department of Biostatistics, Harvard S.T. Chan School of Public Health, 2001-2002, and 2006

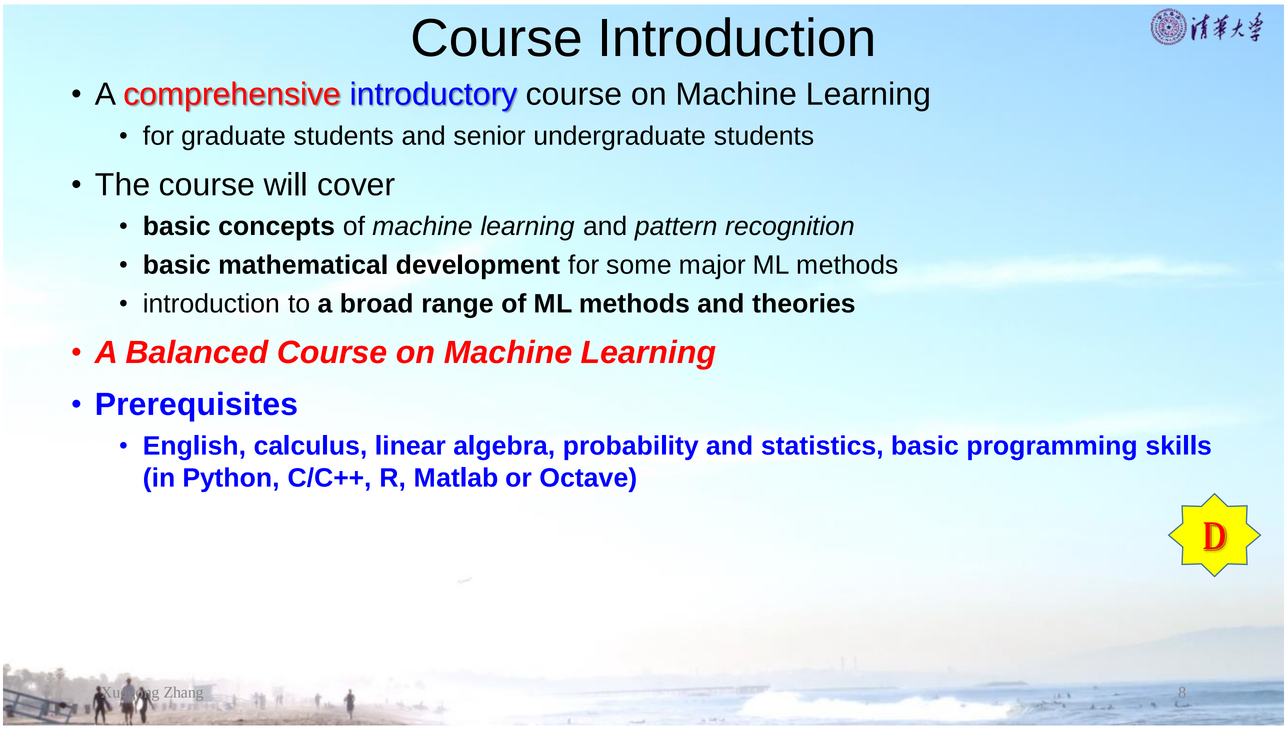
2018
2006
2017
2019

- Please use ~1mins to get familiar with your neighbors, in English



Symbols used in course slides

- Symbols & Meaning
  - R • Extensive reading suggested
  - W • Wikipedia looking-up suggested
  - G • Group discussion encouraged
  - H • Homework point
  - D • In-class discussion



Course Introduction

- A comprehensive introductory course on Machine Learning
  - for graduate students and senior undergraduate students
- The course will cover
  - **basic concepts** of *machine learning* and *pattern recognition*
  - **basic mathematical development** for some major ML methods
  - introduction to **a broad range of ML methods and theories**
- ***A Balanced Course on Machine Learning***
- **Prerequisites**
  - **English, calculus, linear algebra, probability and statistics, basic programming skills (in Python, C/C++, R, Matlab or Octave)**

D

Course Introduction

- It is an EMI course, not an English course.
  - Do not take this course for learning English,
  - but **do use this course to enhance your habit/skill of thinking/studying in English**.
    - Working language in class: English (including Q&As)
    - Homework and exams: answered can be in Chinese or English

D

Composition of the course

- **Lectures**
- In-class discussions
- After-class discussions
- TA office hours (appointment-based)
- **Readings and Reviews**
  - Slides, Reference books, Online materials
- **Homework exercises**
  - Theoretical problems (Pr)
  - Computer exercises (Ex)
- **Mid-term exam**
- **Course project**
- **Final exam**

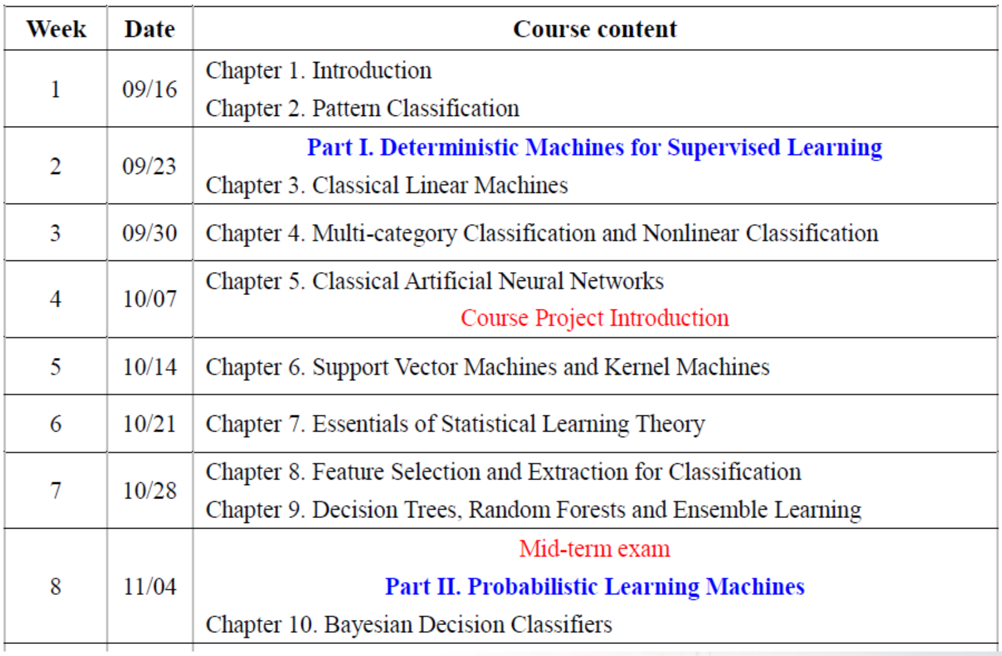
D

Composition of the course

- Grading (up to adjustment)
  - 40% homework (Pr+Ex)
  - 20% course project
  - 5% midterm exam
  - 35 final exam

D

- The honor code:
  - Group learning and reference-checking encouraged, but **homework must be done independently**.
  - Collaborations, discussions, references **must be acknowledged** in the report.
  - **Plagiarism and other misconducts will get negative scores.**



Week	Date	Course content
1	09/16	Chapter 1. Introduction Chapter 2. Pattern Classification
2	09/23	**Part I. Deterministic Machines for Supervised Learning** Chapter 3. Classical Linear Machines
3	09/30	Chapter 4. Multi-category Classification and Nonlinear Classification
4	10/07	Chapter 5. Classical Artificial Neural Networks Course Project Introduction
5	10/14	Chapter 6. Support Vector Machines and Kernel Machines
6	10/21	Chapter 7. Essentials of Statistical Learning Theory
7	10/28	Chapter 8. Feature Selection and Extraction for Classification Chapter 9. Decision Trees, Random Forests and Ensemble Learning
8	11/04	Mid-term exam **Part II. Probabilistic Learning Machines** Chapter 10. Bayesian Decision Classifiers

Syllabus

8	11/04	**Part II. Probabilistic Learning Machines** Chapter 10. Bayesian Decision Classifiers
9	11/11	Chapter 11. Probability Density Estimation Chapter 12. Hidden Markov Models and Introduction to Graphic Models
10	11/18	**Part III. Machines for Unsupervised Learning** Chapter 13. Clustering: Unsupervised Pattern Recognition
11	12/25	Chapter 14. Model-based Clustering and EM Algorithms Chapter 15. Manifold Learning, Dimensionality Reduction and Visualization
12	12/02	Chapter 16. Unsupervised Learning with Neural Networks: SOM and RBM
13	12/09	**Part IV. Deep Learning and Beyond** Chapter 17. Convolution Neural Networks
14	12/16	Chapter 18. Recurrent Neural Networks, LSTM and Transformers
15	12/23	Chapter 19. Deep Neural Networks and Generative Models Chapter 20. Ethics Issues in Machine Learning and AI
16		Final Exam
17		Course Project Report Due

Syllabus (cont'd)

Reference Books

- **张学工、汪小我《模式识别（第四版）：模式识别与机器学习》，清华大学出版社，2021.8**
- R.O. Duda P.E. Hart, D.G. Stork, *Pattern Classification* (2^{nd} edition), John Wiley & Sons, Inc, 2001 (Chinese translation 《模式分类》 also available)
- S. Raschka & V. Mirjalili, *Python Machine Learning* (2nd edition), Birmingham, Packt Publishing, 2017
- Y. S. Abu-Mostafa, M. Magdon-Ismail, H-T. Lin, *Learning from Data*, AMLbook.com, 2012
- Christopher M. Bishop, *Pattern Recognition and Machine Learning*, NY: Springer, 2006
- T. Hastie, R. Tibshirani, J. Friedman, *The Elements of Statistical Learning* (2nd edition), Springer, 2016
- S. Shalev-Shwartz & S. Ben-David, *Understanding Machine Learning: From Theory to Algorithms*, Cambridge University Press, 2014
- I. Goodfellow, Y. Bengio, A. Courville, *Deep Learning*, MIT Press, 2017
- Xuegong Zhang, *Learning Machines: a Balanced Book on Machine Learning*, to be expected

模式识别
模式识别与机器学习
（第4版）
PATTERN RECOGNITION

模式识别
Pattern Recognition
(Third Edition)

Lecture Slides

- Available on our course homepage in the Web Learning system
- For students of this course only

- Designed to assist teaching and after-class review
- Not designed for self-study
  - You may find the slides hard to fellow if you do not attend the lectures.

Course Projects

- To be assigned in Oct.
- The project can be completed in groups
  - Each group should be of 1~3 students
  - Members of the same group may get different scores
  - Group-forming and project selection in mid-Oct.
- Project report should include
  - A formal technical report (ready for journal submission)
  - Supplementary materials: codes/scripts, technical details, all experiment results, other materials that are necessary for reproducing the work

about the homework

- Problem sets: one week time, deadline at **23:00pm Wed**
- Computer exercises: two weeks time, deadline at **23:00pm Wed**
- Later submissions:
  - To be sent to TAs via emails, at the cost of **a 10% daily deduction of the score**, unless you have got **written permission** before the deadline.
  - e.g., If one exercise is 10 points and your completion earns 8, you will get

    $8 \times (1 - 0.1 \times 3) = 5.6$ points if submitted 3 days later, and

    $8 \times (1 - 0.1 \times 7) = 2.4$ points if submitted 7 days later.

D

Any questions?

D

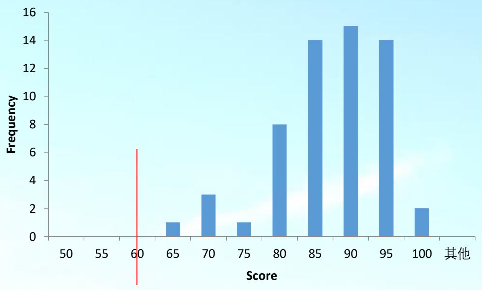
Can I take this course if my math/coding background is weak?

Please quit
if you don't know the meaning
of the underlined words.

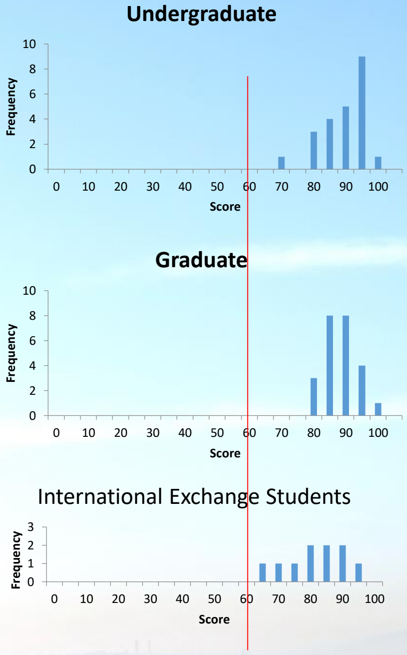
- Prerequisites
  - English
  - Calculus: comfortable with **functions**, **derivatives** and **integrals**
  - Linear algebra: comfortable with **vectors** and **matrixes**
  - Probability and statistics: comfortable with **random variables** and **probability distributions**
  - Basic programming skills: comfortable with **coding** in any computer language, such as Python, C/C++, R, Matlab or Octave

D

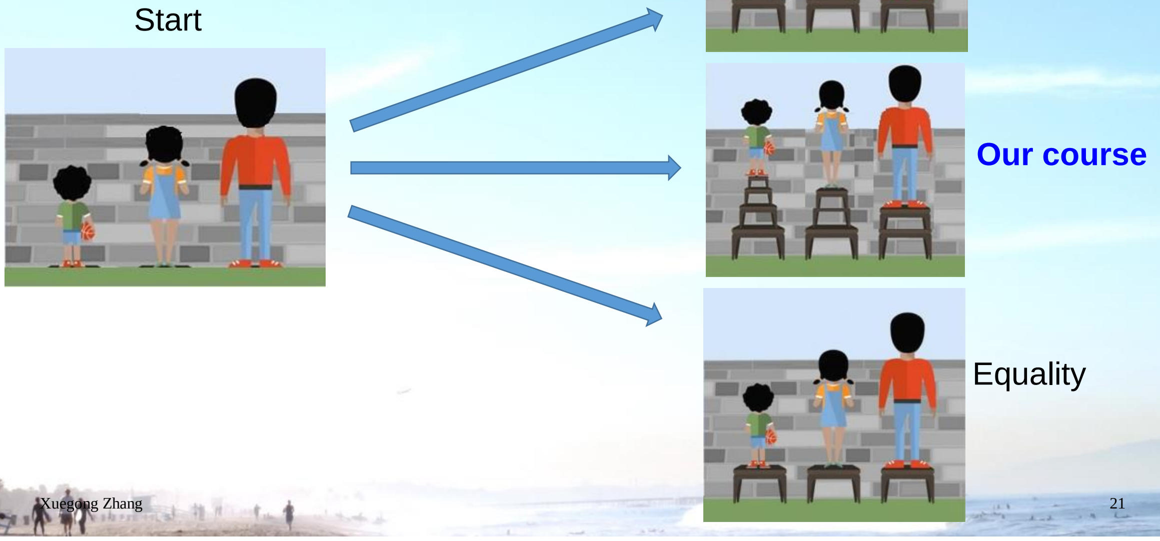
Score distribution in 2019



等级制成绩总人数	58 人，	
A+	0 人，	占 0.0 %；
A	3 人，	占 5.2 %；
A−	16 人，	占 27.6 %；
B+	15 人，	占 25.9 %；
B	13 人，	占 22.4 %；
B−	5 人，	占 8.6 %；
C+	1 人，	占 1.7 %；
C	0 人，	占 0.0 %；
C−	4 人，	占 6.9 %；
D+	0 人，	占 0.0 %；
D	1 人，	占 1.7 %；
F	0 人，	占 0.0 %；



Ideal: a course for everyone



Start

Equity

Our course

Equality

---

stretch break
(1 minute)

D

What is Machine Learning?

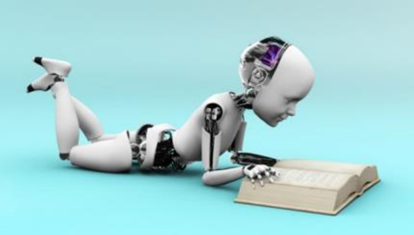
What is “learning”

- Learning is the process of acquiring new, or modifying existing, knowledge, behaviors, skills, values, or preferences.

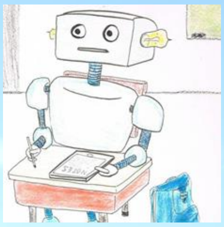
---- Wikipedia

- What do we learn from?
  - From knowledge archives
  - From observations (data)

What is “machine learning”

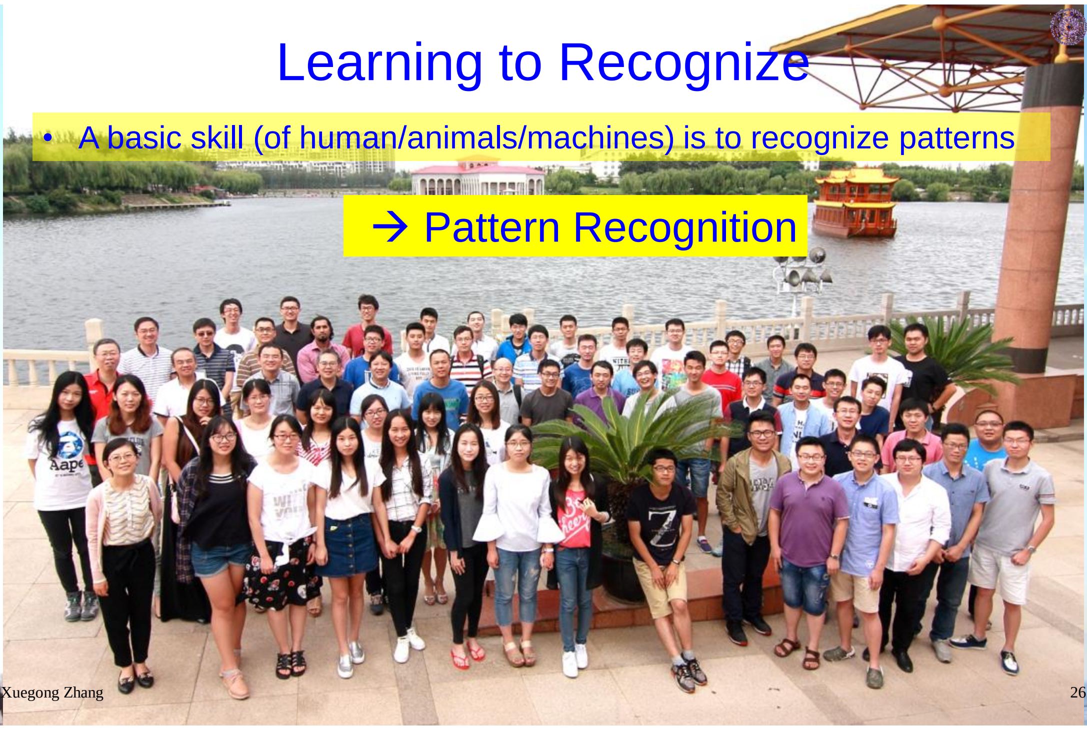


- Knowledge-based learning
- ✓ Data-based learning D



- Learning is the process of acquiring new, or modifying existing, knowledge, behaviors, **skills**, values, or preferences.

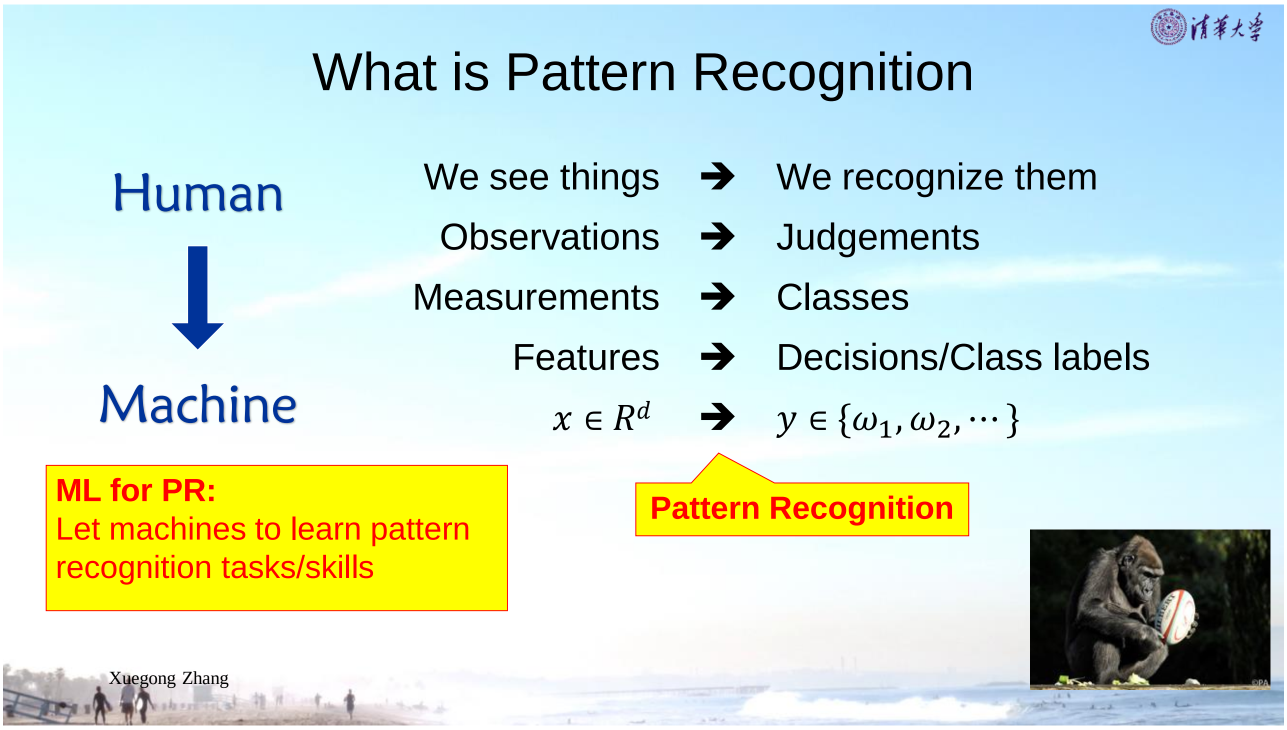
Learning to Recognize



- A basic skill (of human/animals/machines) is to recognize patterns

→ Pattern Recognition

- We recognize objects by their patterns.



What is Pattern Recognition

Human

Machine

We see things ➔ We recognize them

Observations ➔ Judgements

Measurements ➔ Classes

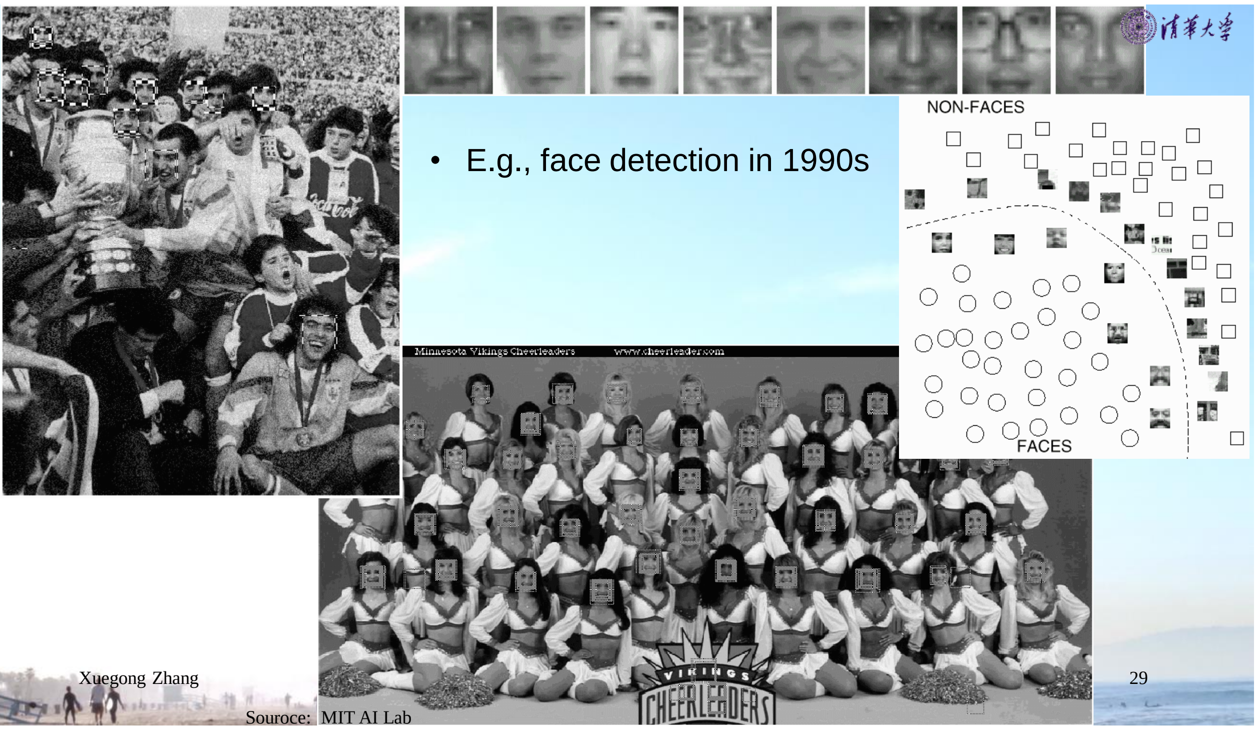
Features ➔ Decisions/Class labels

$x \in R^d$ ➔ $y \in \{\omega_1, \omega_2, \cdots\}$

Pattern Recognition

ML for PR:
Let machines to learn pattern recognition tasks/skills

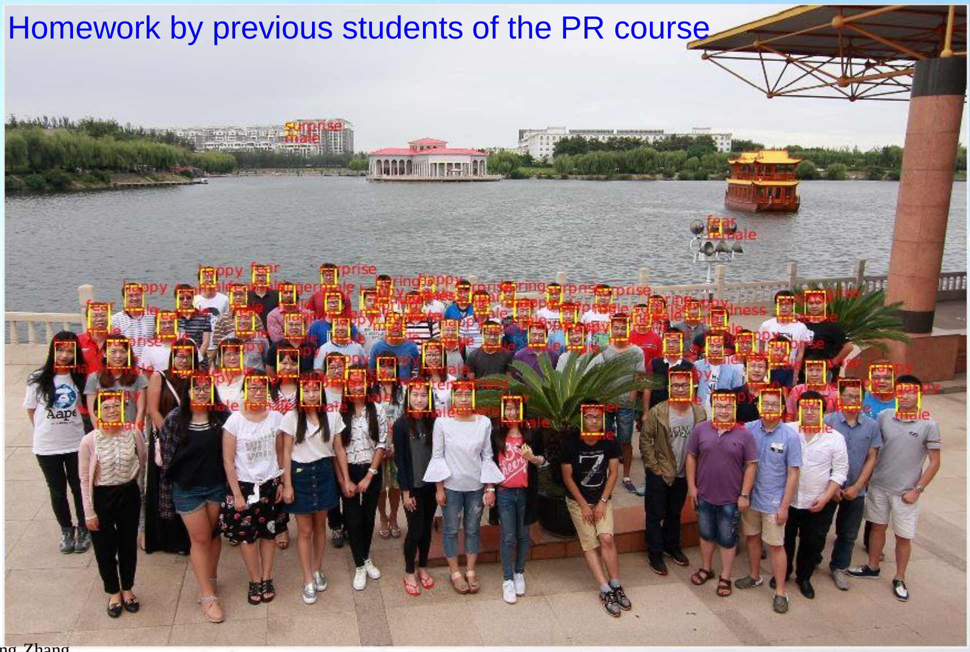
- E.g., face detection in 1990s


NON-FACES
FACES
Minnesota Vikings Cheerleaders
www.cheerleader.com
VIKINGS
CHEERLEADERS

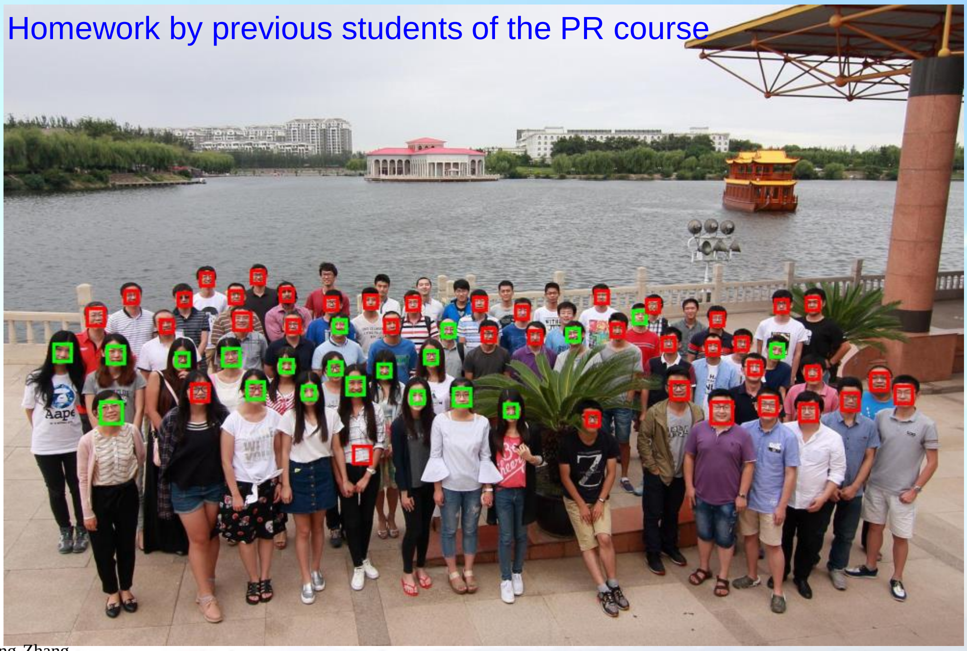
Souroce: MIT AI Lab

E.g. Homework by previous students of the PR course

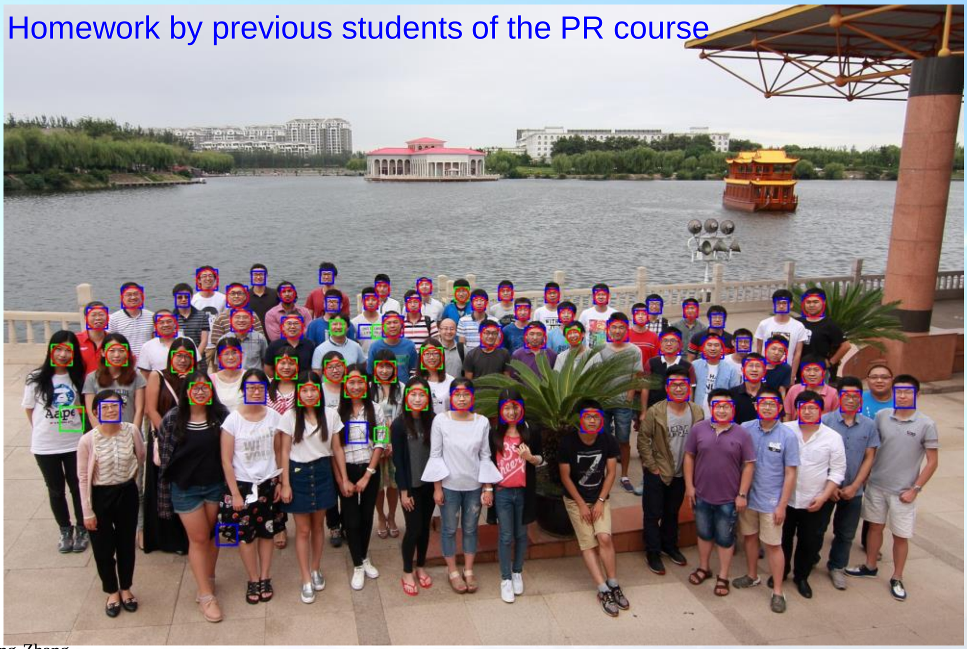
E.g. Homework by previous students of the PR course



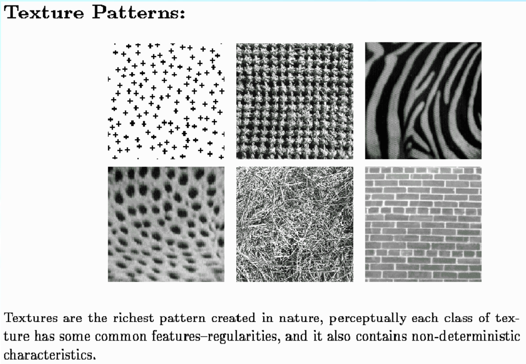
E.g. Homework by previous students of the PR course



E.g. Homework by previous students of the PR course

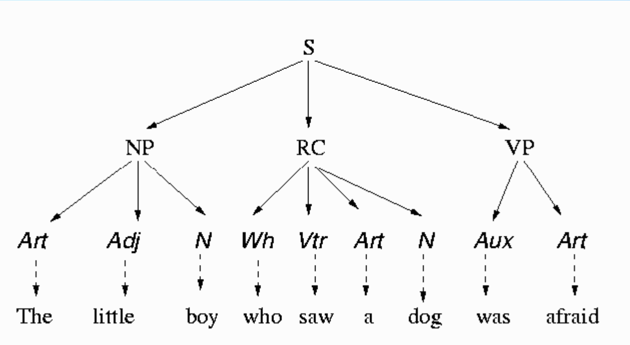


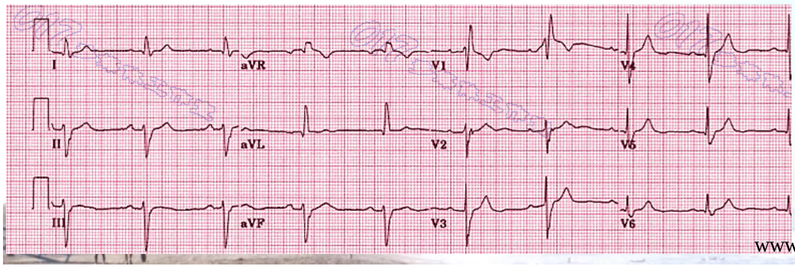
Everything has patterns

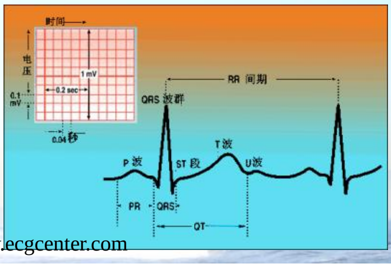


Texture Patterns:

Textures are the richest pattern created in nature, perceptually each class of texture has some common features–regularities, and it also contains non-deterministic characteristics.

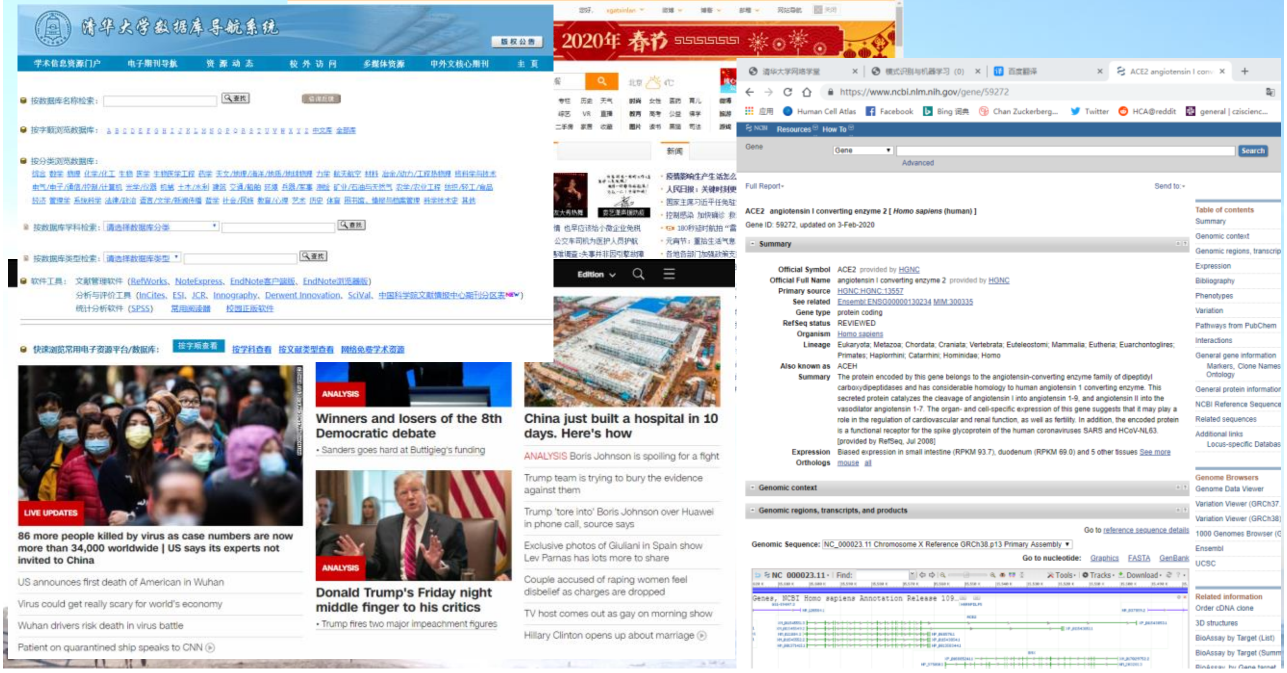

S
NP
RC
VP
Art
Adj
N
Wh
Vtr
Art
N
Aux
Art
The
little
boy
who
saw
a
dog
was
afraid


I
aVR
V1
V4
II
aVL
V2
V5
III
aVF
V3
V6


时间
电压
1 mV
0.1 mV
0.2 sec
0.04秒
RR 间期
QRS 波群
T波
P波
ST 段
U波
PR
QRS
QT

www.ecgcenter.com

Everything has patterns



软件工具：文献管理软件（RefWorks、NoteExpress、EndNote客户端版、EndNote浏览器版）
分析与评价工具（InCites、ESI、JCR、Innography、Derwent Innovation、SciVal、中国科学院文献情报中心期刊分区表）
统计分析软件（SPSS） 常用阅读器 校园正版软件

快速浏览常用电子资源平台/数据库：按字顺查看 按学科查看 按文献类型查看 网络免费学术资源

LIVE UPDATES
86 more people killed by virus as case numbers are now more than 34,000 worldwide | US says its experts not invited to China
US announces first death of American in Wuhan
Virus could get really scary for world's economy
Wuhan drivers risk death in virus battle
Patient on quarantined ship speaks to CNN

ANALYSIS
Winners and losers of the 8th Democratic debate
- Sanders goes hard at Buttigieg's funding

ANALYSIS
Donald Trump's Friday night middle finger to his critics
- Trump fires two major impeachment figures

China just built a hospital in 10 days. Here's how
ANALYSIS Boris Johnson is spoiling for a fight
Trump team is trying to bury the evidence against them
Trump 'tore into' Boris Johnson over Huawei in phone call, source says
Exclusive photos of Giuliani in Spain show Lev Parnas has lots more to share
Couple accused of raping women feel disbelief as charges are dropped
TV host comes out as gay on morning show
Hillary Clinton opens up about marriage

ACE2 angiotensin I converting enzyme 2 [*Homo sapiens* (human)]
Gene ID: 59272, updated on 3-Feb-2020

Summary

Official Symbol	ACE2 provided by HGNC
Official Full Name	angiotensin I converting enzyme 2 provided by HGNC
Primary source	HGNC:HGNC:13557
See related	Ensembl:ENSG00000130234 MIM:300335
Gene type	protein coding
RefSeq status	REVIEWED
Organism	Homo sapiens
Lineage	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo
Also known as	ACEH
Summary	The protein encoded by this gene belongs to the angiotensin-converting enzyme family of dipeptidyl carboxydipeptidases and has considerable homology to human angiotensin 1 converting enzyme. This secreted protein catalyzes the cleavage of angiotensin I into angiotensin 1-9, and angiotensin II into the vasodilator angiotensin 1-7. The organ- and cell-specific expression of this gene suggests that it may play a role in the regulation of cardiovascular and renal function, as well as fertility. In addition, the encoded protein is a functional receptor for the spike glycoprotein of the human coronaviruses SARS and HCoV-NL63. [provided by RefSeq, Jul 2008]
Expression	Biased expression in small intestine (RPKM 93.7), duodenum (RPKM 69.0) and 5 other tissues See more
Orthologs	mouse all

Genomic context

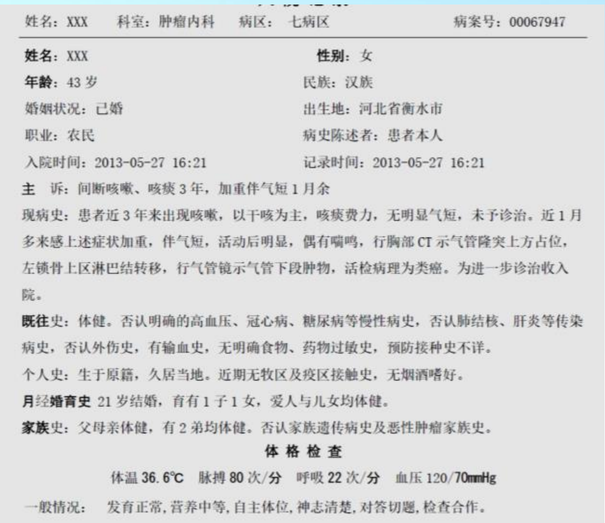
Genomic regions, transcripts, and products

Go to reference sequence details

Genomic Sequence: NC_000023.11 Chromosome X Reference GRCh38.p13 Primary Assembly

Go to nucleotide: Graphics FASTA GenBank

Everything has patterns



姓名：XXX　科室：肿瘤内科　病区：七病区	病案号：00067947
姓名：XXX	**性别**：女
年龄：43 岁	民族：汉族
婚姻状况：已婚	出生地：河北省衡水市
职业：农民	病史陈述者：患者本人
入院时间：2013-05-27 16:21	记录时间：2013-05-27 16:21

主　诉：间断咳嗽、咳痰 3 年，加重伴气短 1 月余

现病史：患者近 3 年来出现咳嗽，以干咳为主，咳痰费力，无明显气短，未予诊治。近 1 月多来感上述症状加重，伴气短，活动后明显，偶有喘鸣，行胸部 CT 示气管隆突上方占位，左锁骨上区淋巴结转移，行气管镜示气管下段肿物，活检病理为类癌。为进一步诊治收入院。

既往史：体健。否认明确的高血压、冠心病、糖尿病等慢性病史，否认肺结核、肝炎等传染病史，否认外伤史，有输血史，无明确食物、药物过敏史，预防接种史不详。

个人史：生于原籍，久居当地。近期无牧区及疫区接触史，无烟酒嗜好。

月经**婚育史** 21 岁结婚，育有 1 子 1 女，爱人与儿女均体健。

家族史：父母亲体健，有 2 弟均体健。否认家族遗传病史及恶性肿瘤家族史。

体 格 检 查

体温 36.6℃　脉搏 80 次/分　呼吸 22 次/分　血压 120/70mmHg

一般情况：　发育正常，营养中等，自主体位，神志清楚，对答切题，检查合作。

主诉：全身皮肤散在瘀点2天。

现病史：患者2天前无明显诱因出现全身皮肤散在瘀点、瘀斑，以四肢为甚，伴有牙龈出血，无血便、血尿及呕血，无畏寒、发热，无咳嗽。未就诊及治疗。今天仍有出现新鲜瘀点，为进一步检查、治疗拟收住院。四天前有腹泻。自发病以来，无畏寒、发热，无头痛、头晕，无咳嗽、咳痰，无胸闷、胸痛及气促。无腹胀、腹痛。大、小便正常，无血尿、血便。食欲、睡眠尚正常。双下肢无浮肿。

既往史：10年前有类似病史，经治疗后好转，具体诊断及治疗不详。否认有高血压病、糖尿病、肝炎、肺结核、血友病等病史。否认有外伤史、手术史及输血史。否认有药物及食物过敏史。预防接种史不详。

个人史：原籍出生、长大，近五年在ZZ工作，从事化妆品制造工作多年。未涉及疫水及传染病区。无嗜酒史。吸烟史6年，10支/天。

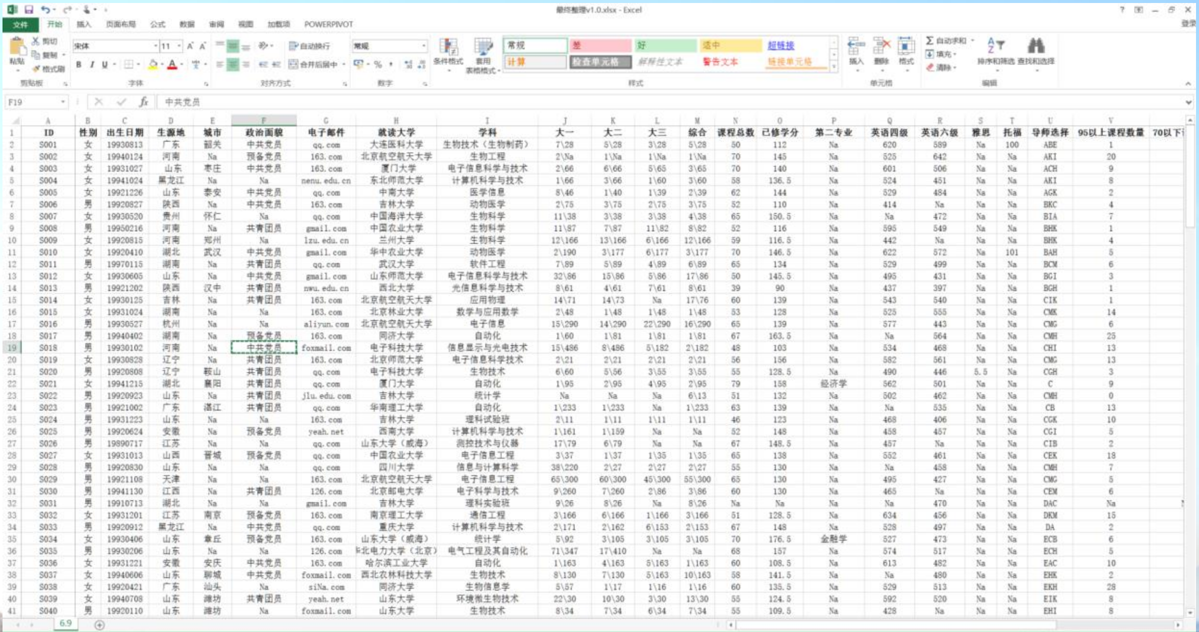
婚姻生育史:未婚、未育。

家族史：否认家族中有类似疾病患者，否认家族中有肝炎、肺结核、高血压病、糖尿病、血友病及肿瘤等疾病。

体格检查：T 36.5℃　P 80次/分　R 20次/分　BP 110/70mmHg
发育正常，营养不良，……

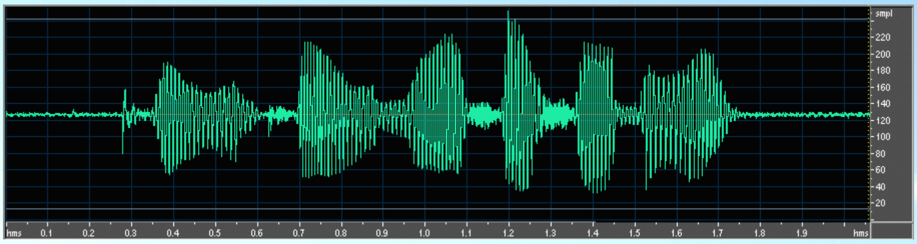
辅助检查
BLOOD RT：WBC：12.1×10*9/L,
HGB:118g/L ,RBC:5.25×10*12/L ,PLT:3×10*9/L

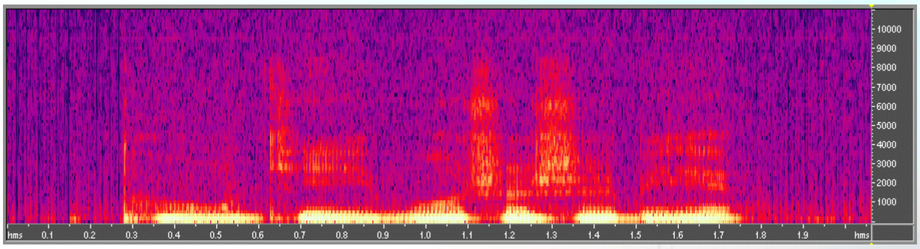
Everything has patterns



ID	性别	出生日期	生源地	城市	政治面貌	电子邮件	就读大学	学科	大一	大二	大三	综合	课程总数	已修学分	第二专业	英语四级	英语六级	雅思	托福	导师选择	95以上课程数量	70以下
S001	女	19930813	广东	韶关	中共党员	qq.com	大连医科大学	生物技术（生物制药）	7\28	5\28	3\28	5\28	50	112	Na	620	589	Na	100	ABE	1	
S002	女	19940124	河南	Na	预备党员	163.com	北京航空航天大学	生物工程	2\Na	1\Na	1\Na	1\Na	70	145	Na	525	642	Na	Na	AKI	20	
S003	女	19931027	山东	枣庄	中共党员	163.com	厦门大学	电子信息科学与技术	2\66	6\66	5\65	3\65	70	140	Na	601	506	Na	Na	ACH	9	
S004	女	19941024	黑龙江	Na	Na	nenu.edu.cn	东北师范大学	计算机科学与技术	1\66	3\66	1\60	3\60	58	136.5	Na	524	451	Na	Na	AKI	8	
S005	女	19921226	山东	泰安	中共党员	qq.com	中南大学	医学信息	8\46	1\40	1\39	2\39	62	144	Na	529	484	Na	Na	AGK	2	
S006	男	19920827	陕西	Na	中共党员	163.com	吉林大学	动物医学	2\75	3\75	2\75	3\75	52	110	Na	414	Na	Na	Na	BKC	4	
S007	女	19930520	贵州	怀仁	Na	qq.com	中国海洋大学	生物科学	11\38	3\38	3\38	4\38	65	150.5	Na	Na	472	Na	Na	BIA	7	
S008	男	19950216	河南	Na	共青团员	gmail.com	中国农业大学	生物科学	11\87	7\87	11\82	8\82	52	116	Na	595	549	Na	Na	BHK	1	
S009	女	19920815	河南	郑州	Na	lzu.edu.cn	兰州大学	生物科学	12\166	13\166	6\166	12\166	59	116.5	Na	442	Na	Na	Na	BHK	4	
S010	女	19920410	湖北	武汉	中共党员	gmail.com	华中农业大学	动物医学	2\190	3\177	6\177	3\177	70	146.5	Na	622	572	Na	101	BAH	5	
S011	男	19970115	湖南	Na	共青团员	qq.com	武汉大学	软件工程	7\89	5\89	4\89	6\89	65	134	Na	529	499	Na	Na	BCM	6	
S012	女	19930605	山东	Na	中共党员	gmail.com	山东师范大学	电子信息科学与技术	32\86	15\86	5\86	17\86	50	145.5	Na	495	431	Na	Na	BGI	3	
S013	男	19921202	陕西	汉中	共青团员	nwu.edu.cn	西北大学	光信息科学与技术	8\61	4\61	7\61	8\61	39	90	Na	437	397	Na	Na	BGH	1	
S014	女	19930125	吉林	Na	共青团员	163.com	北京航空航天大学	应用物理	14\71	14\73	Na	17\76	60	139	Na	543	540	Na	Na	CIK	1	
S015	女	19931024	湖南	Na	Na	163.com	北京林业大学	数学与应用数学	2\48	1\48	1\48	1\48	53	128	Na	525	555	Na	Na	CMK	14	
S016	男	19930527	杭州	Na	Na	aliyun.com	北京航空航天大学	电子信息	15\290	14\290	22\290	16\290	65	139	Na	577	443	Na	Na	CMG	6	
S017	男	19940402	湖南	Na	预备党员	163.com	同济大学	自动化	1\60	1\81	1\81	1\81	67	163.5	Na	Na	564	Na	Na	CMH	25	
S018	男	19930102	河南	Na	中共党员	foxmail.com	电子科技大学	信息显示与光电技术	15\486	8\486	5\182	2\182	48	103	Na	534	468	Na	Na	CHI	13	
S019	女	19930828	辽宁	Na	共青团员	163.com	北京师范大学	电子信息科学技术	2\21	2\21	2\21	2\21	56	156	Na	582	561	Na	Na	CMG	13	
S020	男	19920808	辽宁	鞍山	共青团员	qq.com	电子科技大学	生物技术	6\60	5\56	3\55	3\55	55	128.5	Na	490	446	5.5	Na	CGH	3	
S021	女	19941215	湖北	襄阳	共青团员	qq.com	厦门大学	自动化	1\95	2\95	4\95	2\95	79	158	经济学	562	501	Na	Na	C	9	
S022	男	19920923	山东	Na	共青团员	jlu.edu.com	吉林大学	统计学	Na	Na	Na	6\13	51	132	Na	502	462	Na	Na	CMH	0	
S023	男	19921002	广东	湛江	共青团员	qq.com	华南理工大学	自动化	1\233	1\233	Na	1\233	63	139	Na	Na	535	Na	Na	CB	13	
S024	男	19931223	山东	Na	Na	163.com	吉林大学	理科试验班	2\11	1\11	1\11	1\11	46	123	Na	468	406	Na	Na	CGK	10	
S025	男	19920624	安徽	Na	预备党员	yeah.net	西南大学	计算机科学与技术	1\161	1\159	Na	Na	52	148	Na	458	457	Na	Na	CGI	5	
S026	男	19890717	江苏	Na	Na	qq.com	山东大学（威海）	测控技术与仪器	17\79	6\79	Na	Na	67	148.5	Na	457	Na	Na	Na	CIB	2	
S027	女	19931013	山西	晋城	预备党员	qq.com	中国农业大学	电子信息工程	3\37	1\37	1\35	1\35	65	138	Na	552	461	Na	Na	CEK	18	
S028	男	19920830	山东	Na	Na	qq.com	四川大学	信息与计算科学	38\220	2\27	2\27	2\27	55	130	Na	Na	458	Na	Na	CMH	7	
S029	男	19921108	天津	Na	Na	163.com	北京航空航天大学	电子信息工程	65\300	60\300	45\300	55\300	65	130	Na	495	427	Na	Na	CMG	5	
S030	男	19941130	江西	Na	共青团员	126.com	北京邮电大学	电子科学与技术	9\260	7\260	2\86	3\86	60	130	Na	465	Na	Na	Na	CEM	6	
S031	男	19910713	湖北	Na	Na	gmail.com	吉林大学	理科实验班	9\26	8\26	Na	8\26	Na	Na	Na	Na	470	Na	Na	DAC	Na	
S032	女	19931201	江苏	南京	预备党员	163.com	南京理工大学	通信工程	3\166	6\166	1\166	3\166	51	128.5	Na	634	456	Na	Na	DKM	15	
S033	男	19920912	黑龙江	Na	中共党员	qq.com	重庆大学	计算机科学与技术	2\171	2\162	6\153	2\153	67	148	Na	528	497	Na	Na	DA	2	
S034	女	19930406	山东	章丘	预备党员	163.com	山东大学（威海）	统计学	5\92	3\105	3\105	3\105	70	176.5	金融学	527	473	Na	Na	ECB	6	
S035	男	19930206	山东	Na	Na	126.com	北电力大学（北京）	电气工程及其自动化	71\347	17\410	Na	Na	68	157	Na	574	517	Na	Na	ECH	5	
S036	女	19931221	安徽	安庆	中共党员	163.com	哈尔滨工业大学	自动化	1\163	4\163	5\163	1\163	60	108.5	Na	613	482	Na	Na	EAC	10	
S037	女	19940606	山东	聊城	中共党员	foxmail.com	西北农林科技大学	生物技术	8\130	7\130	5\163	10\163	58	141.5	Na	Na	480	Na	Na	EHK	2	
S038	女	19920421	广东	汕头	Na	siNa.com	同济大学	生物信息学	5\57	1\17	1\16	1\16	60	135.5	Na	529	513	Na	Na	EKH	28	
S039	女	19940708	山东	潍坊	共青团员	yeah.net	山东大学	环境微生物技术	22\30	10\30	3\30	13\30	55	124.5	Na	592	520	Na	Na	EIK	8	
S040	男	19920110	山东	潍坊	Na	foxmail.com	山东大学	生物技术	8\34	7\34	6\34	7\34	55	109.5	Na	428	Na	Na	Na	EHI	8	

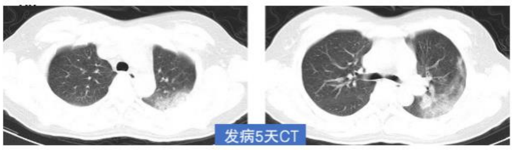
Everything has patterns



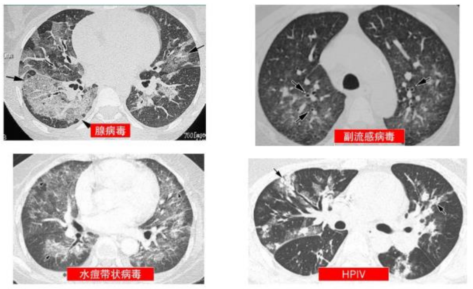


Everything has patterns

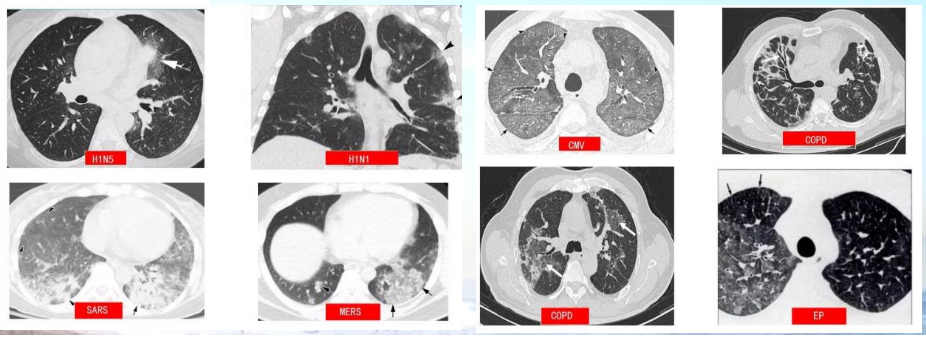
（男性，55 岁，华南海鲜市场商户，发热、咳嗽 10 天）



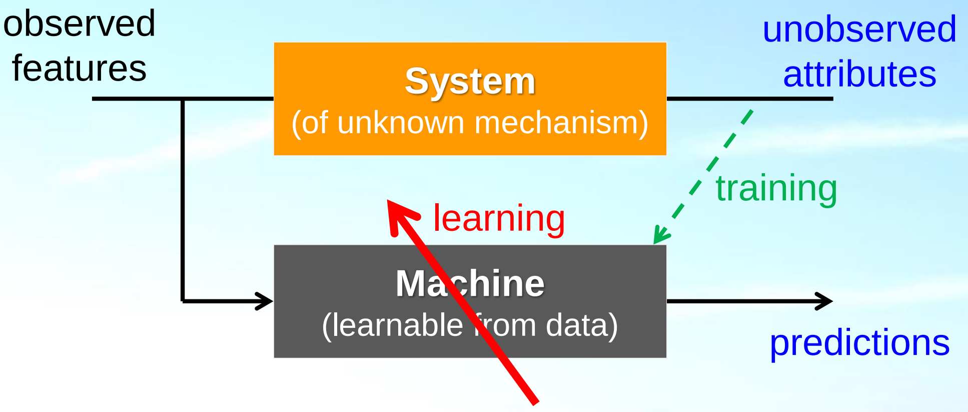
（女性，54 岁，华南海鲜市场商户，发热、气短干咳）


腺病毒
副流感病毒
水痘带状病毒
HPIV

http://infect.dxy.cn/article/676200


H1N5
H1N1
CMV
COPD
SARS
MERS
COPD
EP

Basic Framework of Machine Learning



observed features

System
(of unknown mechanism)

unobserved attributes

training

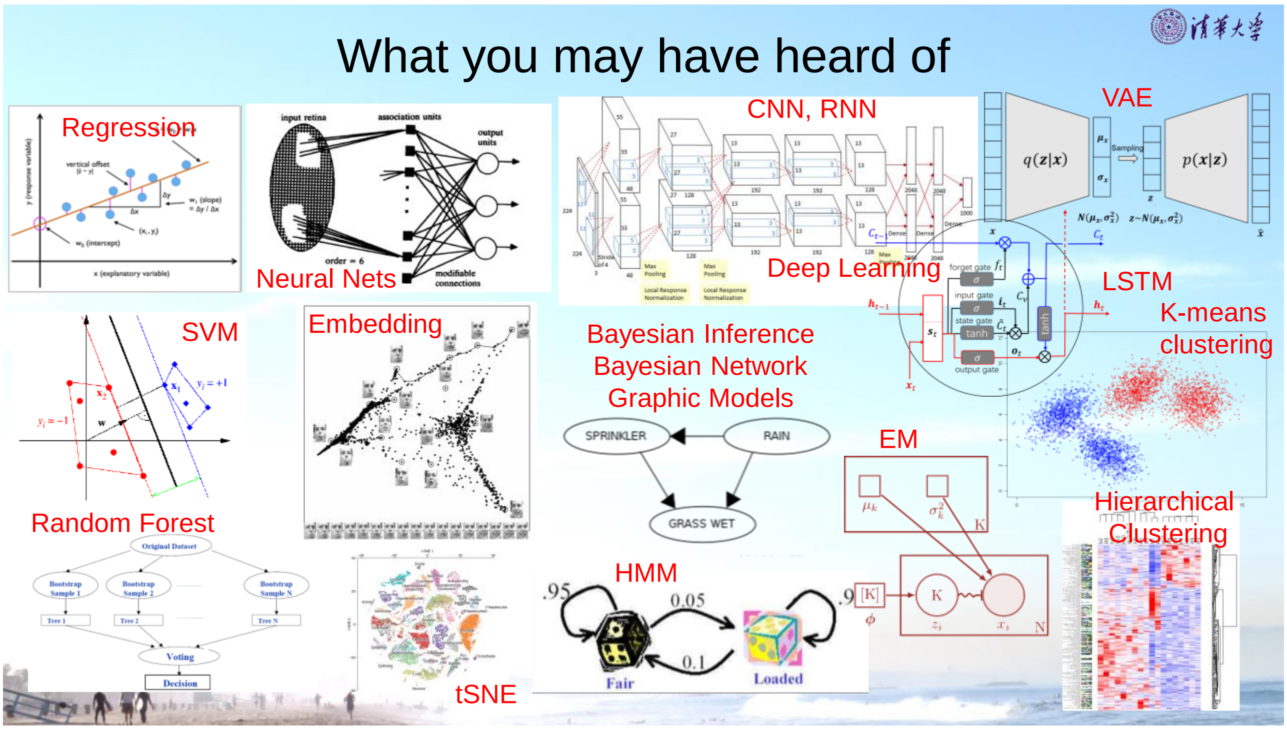
learning

Machine
(learnable from data)

predictions

$$y = f(x)$$

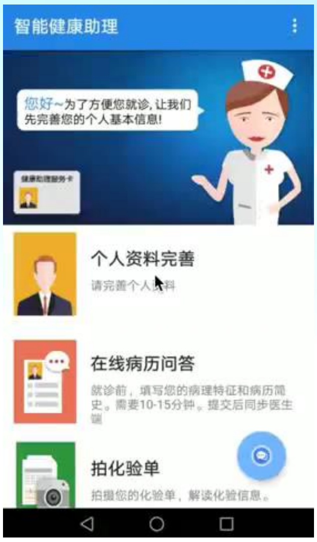
What you may have heard of


Regression
Neural Nets
CNN, RNN
Deep Learning
VAE
$q(z|x)$
$p(x|z)$
LSTM
K-means
clustering
SVM
Embedding
Bayesian Inference
Bayesian Network
Graphic Models
SPRINKLER
RAIN
GRASS WET
EM
Random Forest
Original Dataset
Bootstrap Sample 1
Bootstrap Sample 2
Bootstrap Sample N
Tree 1
Tree 2
Tree N
Voting
Decision
HMM
.95
0.05
0.1
.9
Fair
Loaded
tSNE
Hierarchical
Clustering

What you may have not heard of

- Performance assessment and experiment design
- Connections among major schools of ML methods
- Relations of machine learning, pattern recognition and AI
- Statistical learning theory and regularization
- Learning vs. inference vs. discovery
- Many discussions and personal views
- Ethics of AI and machine learning

Example: ML applications in medicine/health


智能健康助理

您好~为了方便您就诊, 让我们先完善您的个人基本信息!

健康助理服务卡

个人资料完善

请完善个人资料

在线病历问答

就诊前，填写您的病理特征和病历简史。需要10-15分钟。提交后同步医生端

拍化验单

拍摄您的化验单，解读化验信息。

湘雅医院小雅医疗
图像辅助操作系统

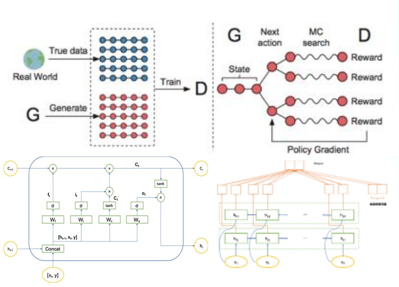
闫海荣
lvhairong@tsinghua.edu.cn

Example: a machine learns to fake medical statements

现病史:3天前无明显原因出现咳嗽，有痰不能自行咳出，伴喘憋，不能平卧，无发热，无腹泻，无便血及黑便，无尿频、尿急及排尿困难，症状进行性加重，今日患者家属发现患者不能自行进食。来我院门诊查血常规白细胞6.4×109/L，中性粒细胞89.6%，血红蛋白103g/l，血小板123×109/L，胸片双肺间质样改变，合并双下肺渗出病变，心电图为窦性心律，异常Q波，ST段改变。为进一步诊治住院治疗。起病后神志清，精神弱，饮食差，睡眠差，二便失禁。
既往史:20余年前患2型糖尿病，目前服用糖适平1片 tid,拜糖平1片 tid,3年前患高血压病，血压最高为180/90mmHg，目前服用硝苯地平,1年前患者不能与人正常交流；1月前出现双下肢肿胀；曾因子宫肌瘤行手术治疗。右眼因白内障手术治疗，左眼失明。否认有肝炎、结核病史，否认有外伤史，否认有食物及药物过敏史。

Training data: real EMR

The machine


True data
Real World
G
Generate
Train
D
G
Next action
MC search
D
State
Reward
Reward
Reward
Reward
Policy Gradient
Concat
tanh
σ
Wf
Wi
Wc
Wo
[xt, y]

关嘉麒

Synthetic EMR “faked” by machine

仿真肺炎病例

1. 患者于 1 周前无明显诱因出现咳嗽，咳白色粘痰，伴活动后加重，休息后可缓解，间断服用镇咳药物等治疗，未行正规诊治。

2. 患者 9 年前受凉后出现咳嗽、咳痰、气喘，活动后加重，喘憋明显，给予抗感染等对症治疗后好转，后进食后症状可改善。

3. 患者 1 年前无明显诱因出现咳嗽、咳痰伴气喘，多于受冷或天气转凉时发作，持续时间在 2 周至半年不等，经抗感染等治疗后好转。无明显诱因，冬季多发。

Pneumonia “patients”

仿真肺癌病例

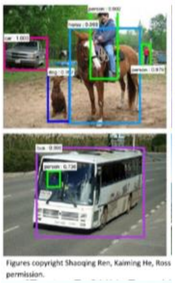
1. 患者反复出现饮食呛咳，无咯血、胸痛、咯血，就诊于北京友谊医院查胸部 CT 示右肺中叶占位性病变，因病变较小定期，于急诊后间断咳嗽、咳少量白痰。

2. 患者 10 多年前开始出现咳嗽、咳痰，痰中带血，当地医院查胸部 CT 示纵隔肿大淋巴结，右肺下叶结节，后于北京人民医院行气管镜时治疗收入院。

3. 患者间断用中药治疗，偶有咯血。4 年前体检症状右侧酸痛，无胸闷、胸痛，无心悸、头晕。2015-9-2 出现发热。

Lung cancer “patients”

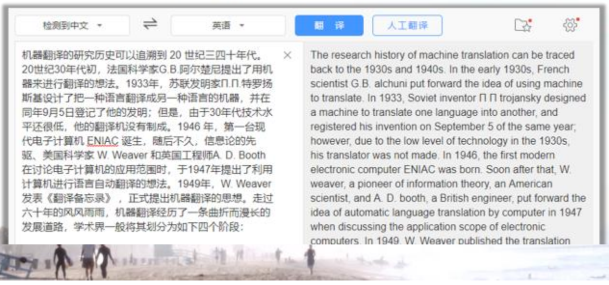
Machines tend to learn everything



Figures copyright Shaoqing Ren, Kaiming He, Ross ... permission.

A white teddy bear sitting in the grass

A man riding a dirt bike on a dirt track

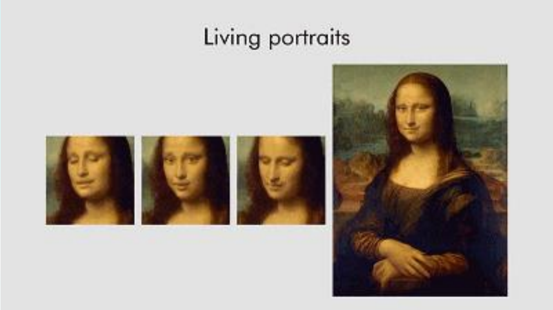


检测到中文
英语
翻 译
人工翻译

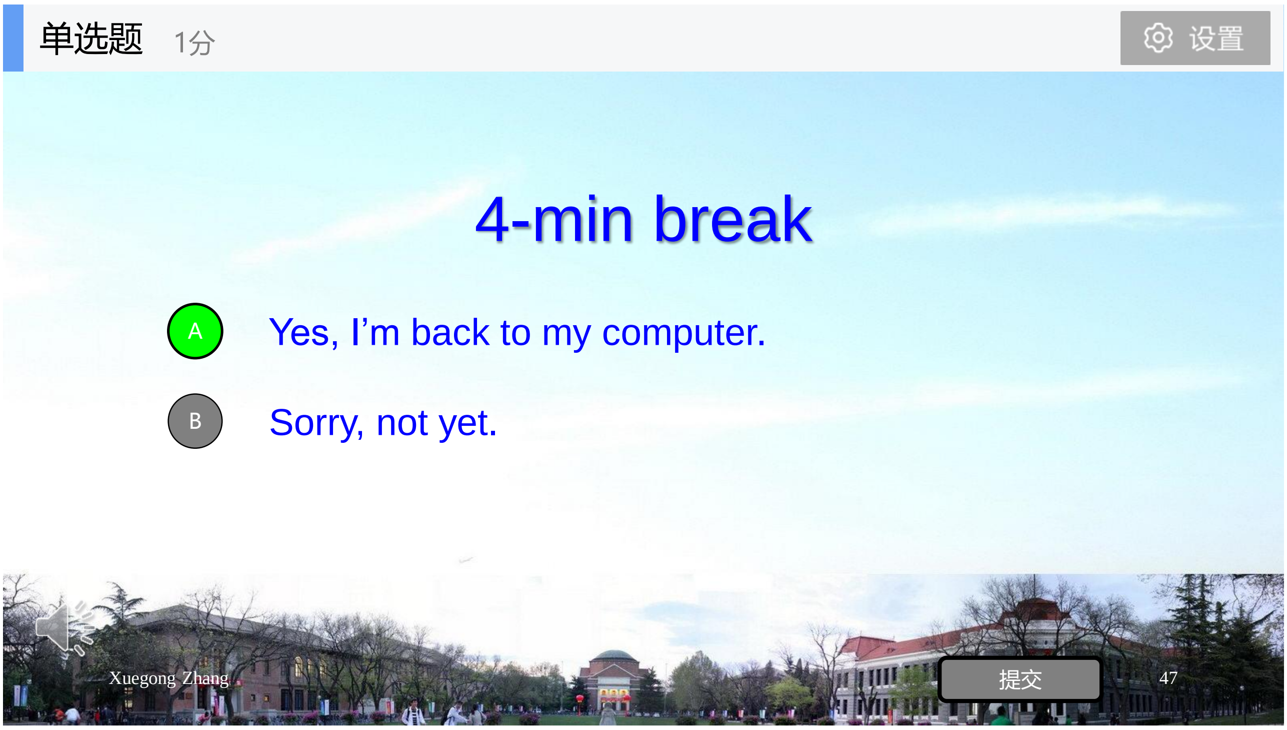
机器翻译的研究历史可以追溯到 20 世纪三四十年代。20世纪30年代初，法国科学家G.B.阿尔楚尼提出了用机器来进行翻译的想法。1933年，苏联发明家П.П.特罗扬斯基设计了把一种语言翻译成另一种语言的机器，并在同年9月5日登记了他的发明；但是，由于30年代技术水平还很低，他的翻译机没有制成。1946 年，第一台现代电子计算机 ENIAC 诞生，随后不久，信息论的先驱、美国科学家 W. Weaver 和英国工程师A. D. Booth 在讨论电子计算机的应用范围时，于1947年提出了利用计算机进行语言自动翻译的想法。1949年，W. Weaver 发表《翻译备忘录》，正式提出机器翻译的思想。走过六十年的风风雨雨，机器翻译经历了一条曲折而漫长的发展道路，学术界一般将其划分为如下四个阶段：

The research history of machine translation can be traced back to the 1930s and 1940s. In the early 1930s, French scientist G.B. alchuni put forward the idea of using machine to translate. In 1933, Soviet inventor П П trojansky designed a machine to translate one language into another, and registered his invention on September 5 of the same year; however, due to the low level of technology in the 1930s, his translator was not made. In 1946, the first modern electronic computer ENIAC was born. Soon after that, W. weaver, a pioneer of information theory, an American scientist, and A. D. booth, a British engineer, put forward the idea of automatic language translation by computer in 1947 when discussing the application scope of electronic computers. In 1949, W. Weaver published the translation

… and do things beyond learning



Living portraits


设置
提交
Xuegong Zhang
47

单选题 1分

4-min break

A Yes, I’m back to my computer.

B Sorry, not yet.