机器学习
前世今生

2017-06-18

About Me

史春奇

简介

- Search Engine – **北京**
- Translation Agent – **京都**
- Language Application Grid – **波士顿**
- Data Analysis – **上海**

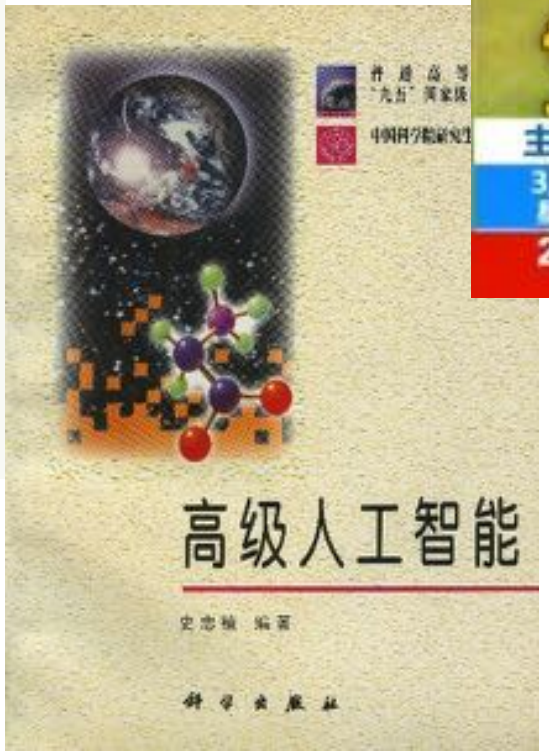
我的导师: 史忠植

人工智能科学技术奖得主史忠植：“智能”人生

史忠植

■ 人物档案

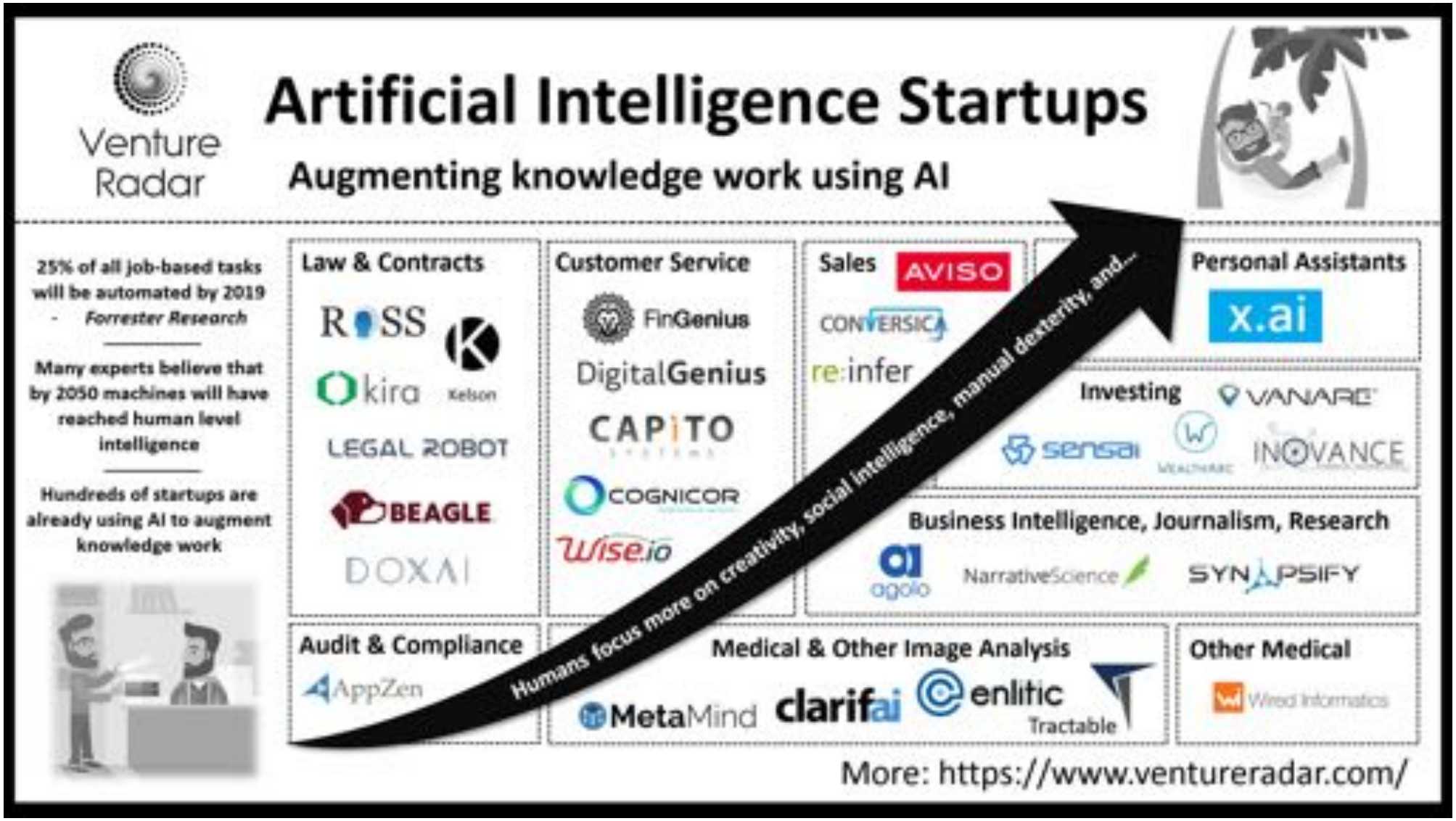
中国科学院计算技术研究所研究员，博士生指导教师，IFIP人工智能专业委员会机器学习和数据挖掘组主席，IEEE高级会员，AAAI和ACM会员，中国人工智能学会和中国计算机学会会士，曾担任中国人工智能学会副理事长、中国计算机学会秘书长。长期从事智能科学、人工智能等方面的研究工作。



大众创业
说起

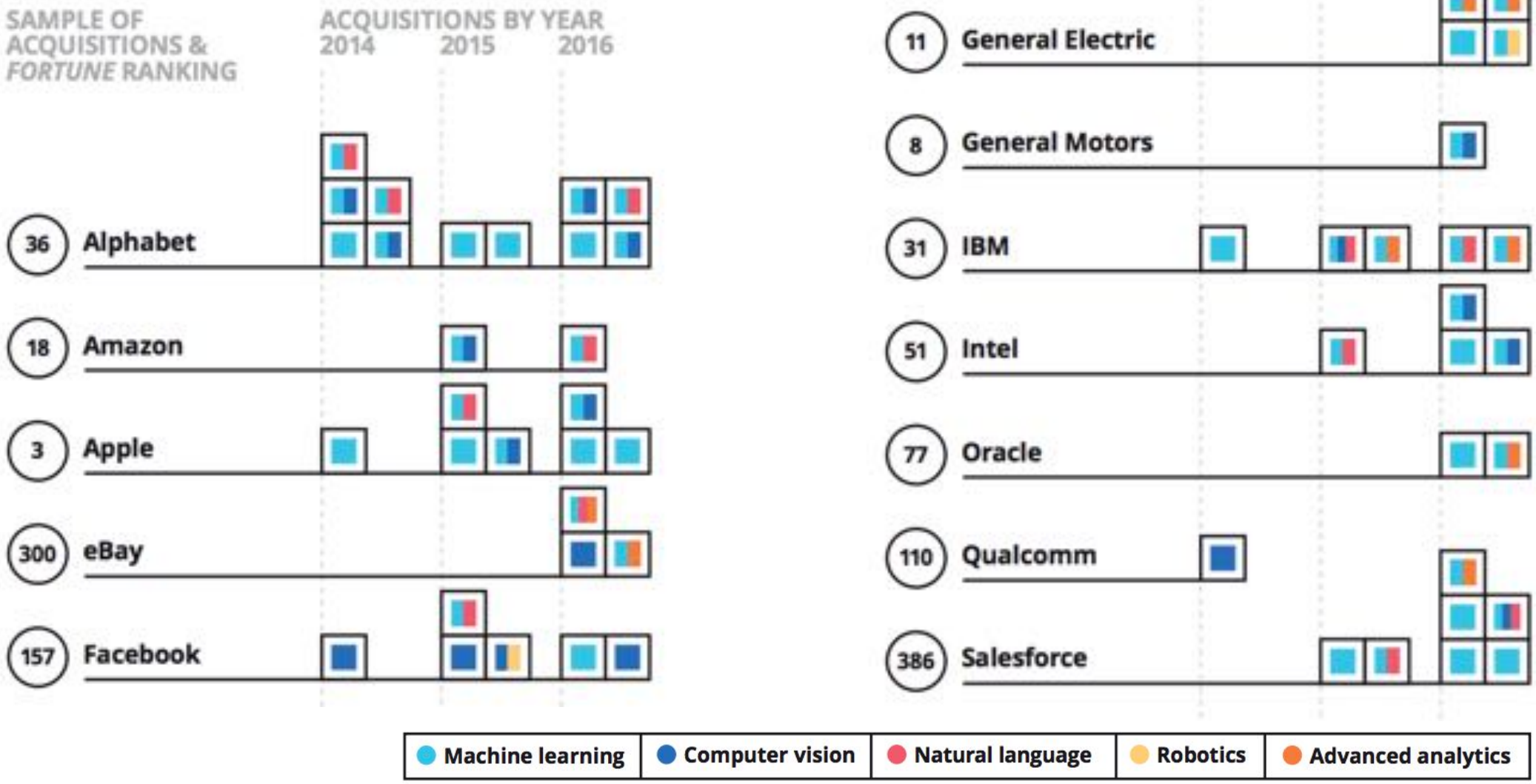
史春奇

初创公司大爆炸


Venture Radar
Artificial Intelligence Startups
Augmenting knowledge work using AI
25% of all job-based tasks will be automated by 2019 - Forrester Research
Many experts believe that by 2050 machines will have reached human level intelligence
Hundreds of startups are already using AI to augment knowledge work
Humans focus more on creativity, social intelligence, manual dexterity, and...
Law & Contracts
ROSS
Kira
Kelson
LEGAL ROBOT
BEAGLE
DOXAI
Customer Service
FinGenius
DigitalGenius
CAPiTO
COGNICOR
Wise.io
Sales
AVISO
CONVERSICA
re:infer
Personal Assistants
x.ai
Investing
VANARE
sensai
WEALTHARC
INOVANCE
Business Intelligence, Journalism, Research
aiaiai
agolo
NarrativeScience
SYNAPSIFY
Audit & Compliance
AppZen
Medical & Other Image Analysis
MetaMind
clarifai
enlitic
Tractable
Other Medical
Wired Informatics
More: https://www.ventureradar.com/

史春奇

国外巨头的收购热潮



国内创业公司


国内人工智能创业公司数量情况
单位（家）
150
100
50
0
92
96
77
26
35
机器人
计算机视觉
自然语言处理
深度学习
智能驾驶

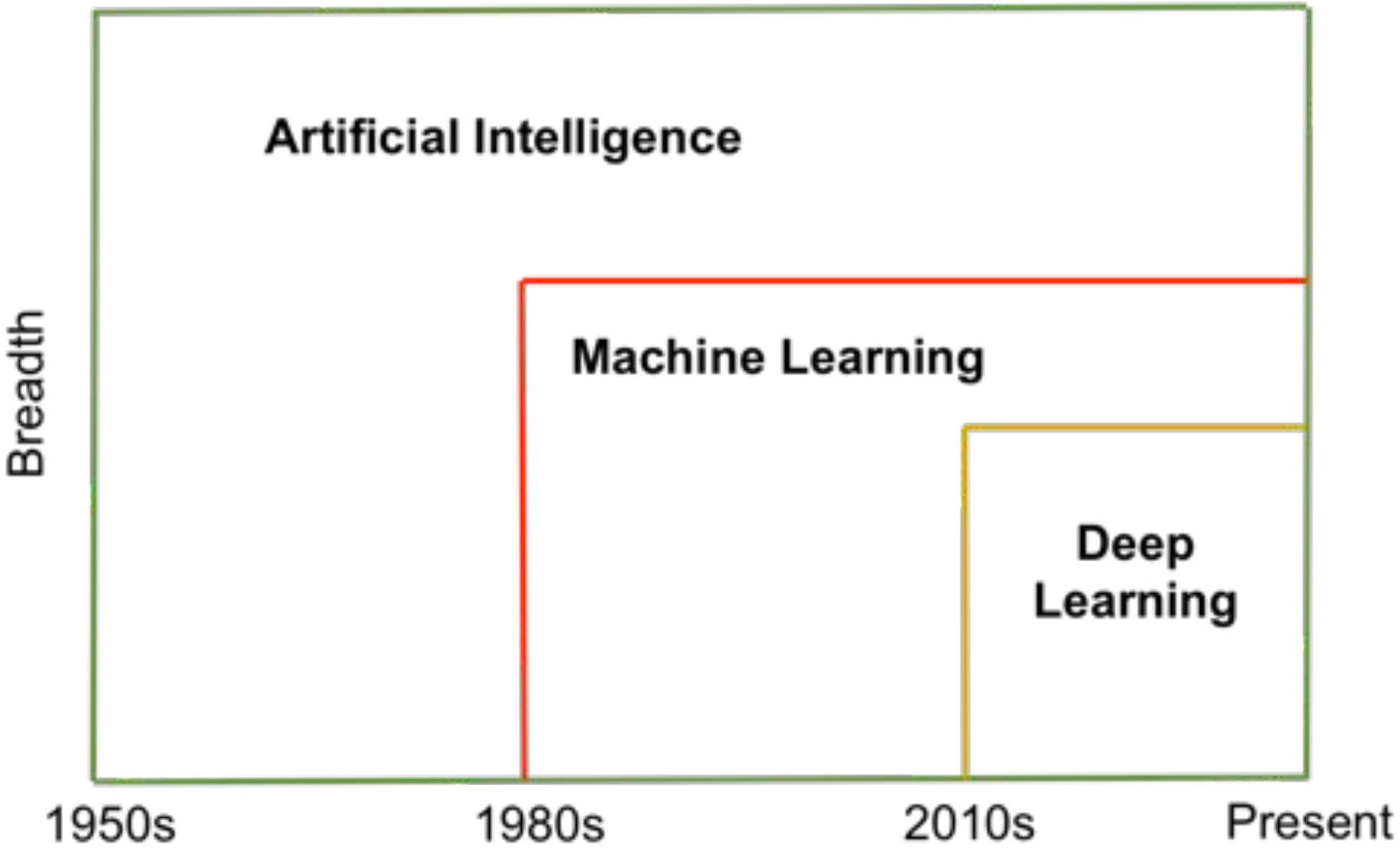

中国人工智能创业公司成立年份占比分布情况
9.50%
7.20%
11.30%
20.80%
34.50%
16.70%
2011年及以前
2012年
2013年
2014年
2015年
2016年


中国人工智能创业公司所属领域分布情况
20.30%
19.60%
10.70%
6.50%
5.90%
5.90%
4.80%
3.60%
计算机视觉
服务机器人
语音及自然语言处理
智能治疗
智能驾驶
机器学习
智能家居
个人助手
智能教育
智能金融
电商物流
垂直应用
数据服务
其他

WHY？

是什么导致这一波创业热潮？

深度学习→机器学习→人工智能

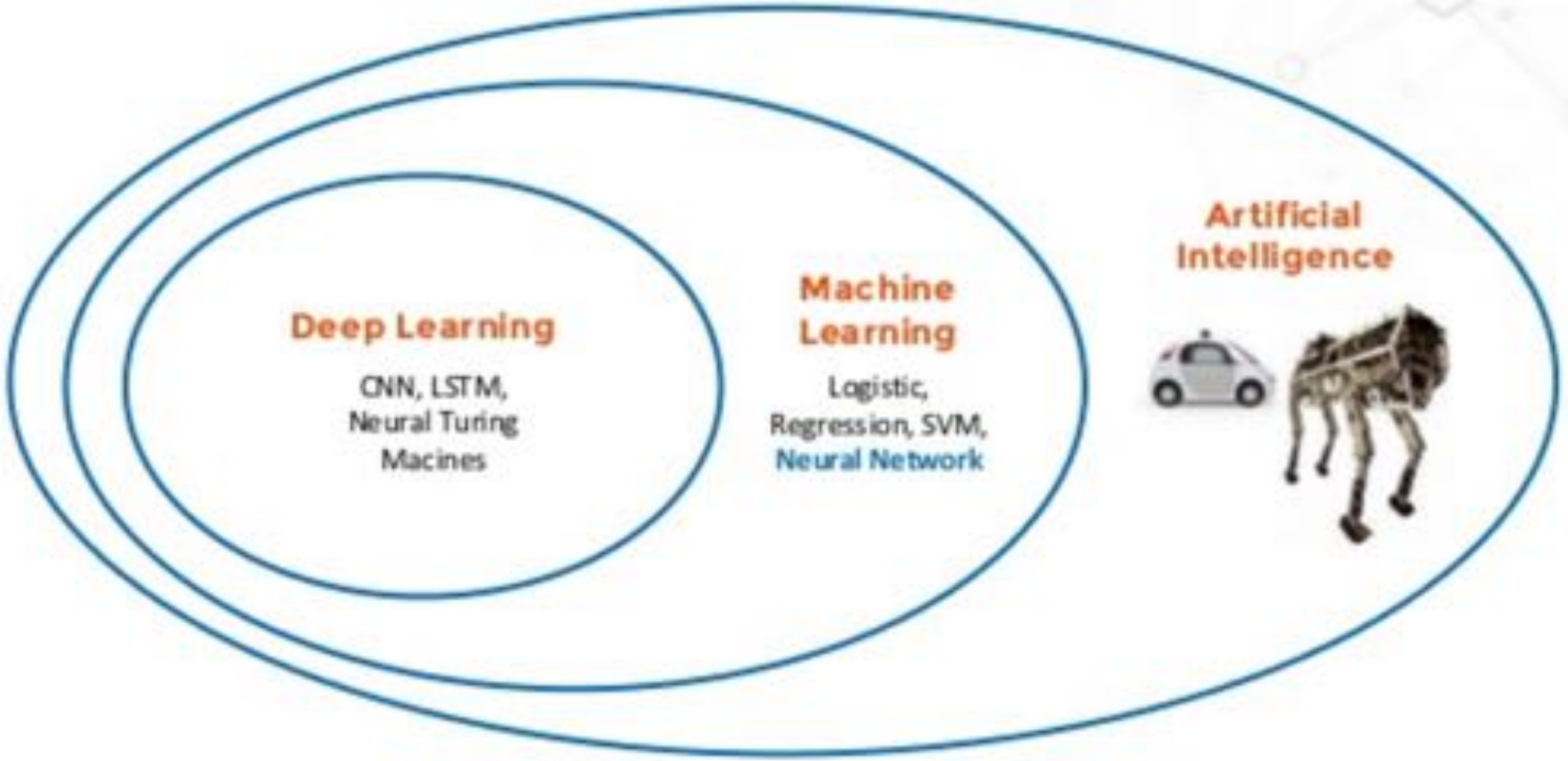


史春奇

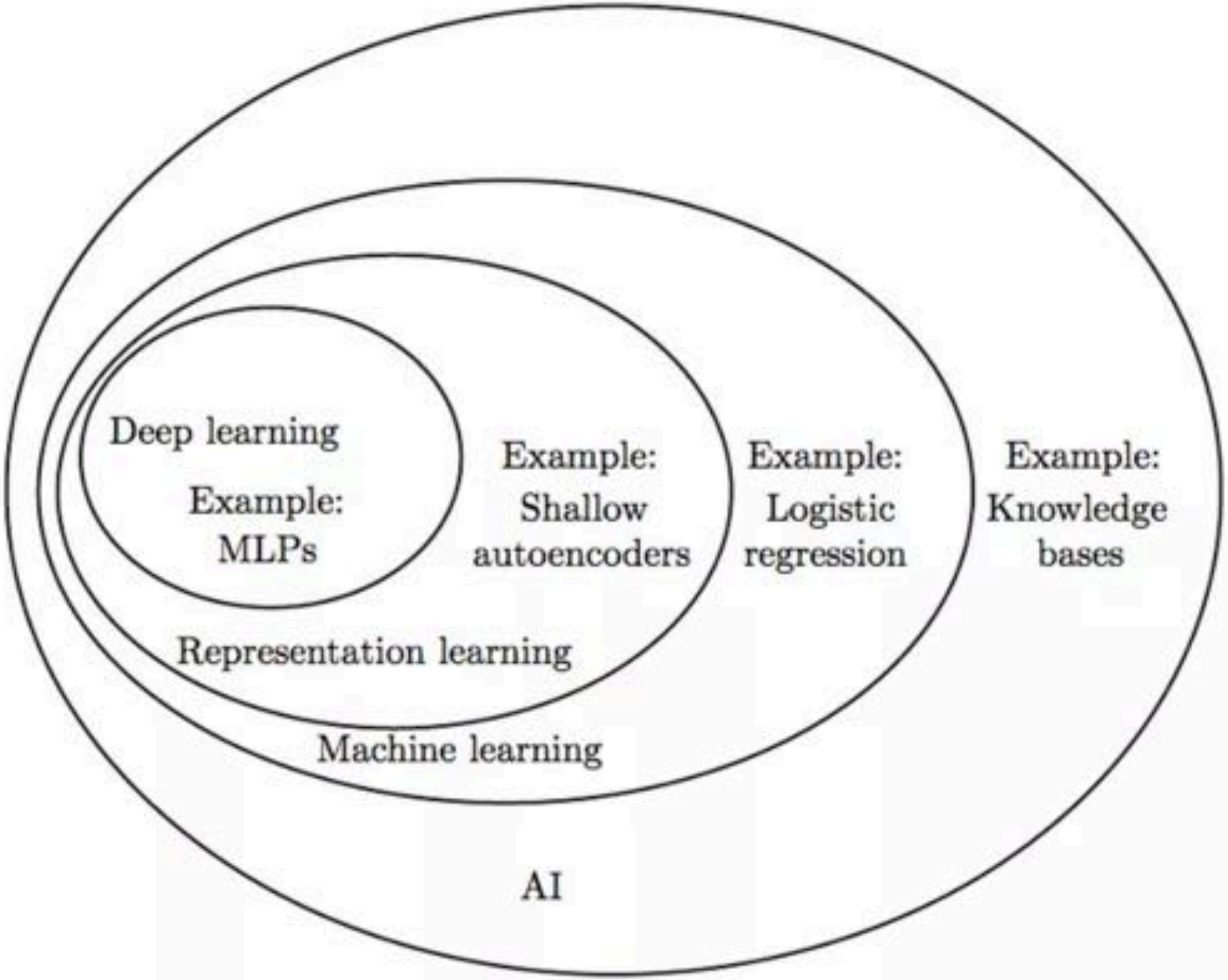
机器学习 和 人工智能

机器学习是目前人工智能中最具希望的部分！

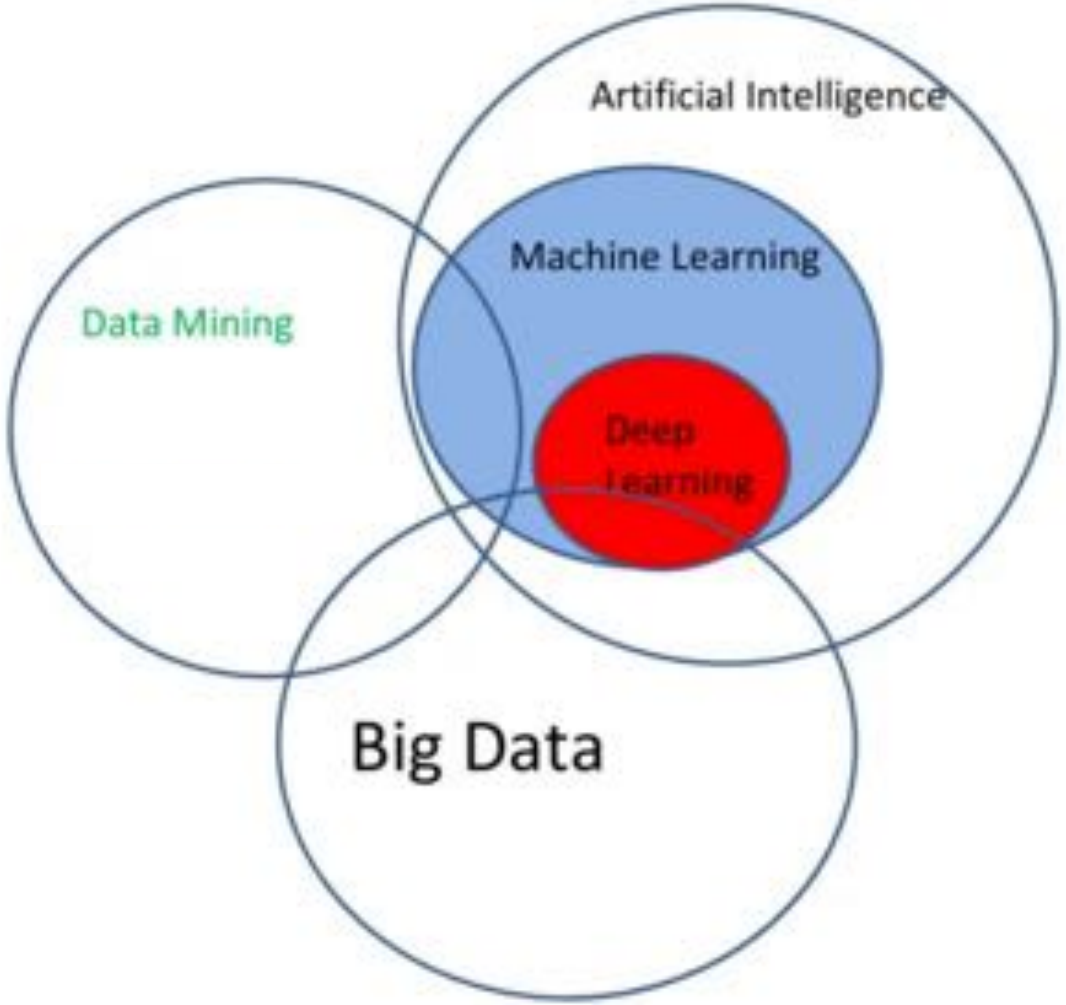
典型例子


Deep Learning
CNN, LSTM,
Neural Turing
Macines
Machine
Learning
Logistic,
Regression, SVM,
Neural Network
Artificial
Intelligence

典型例子

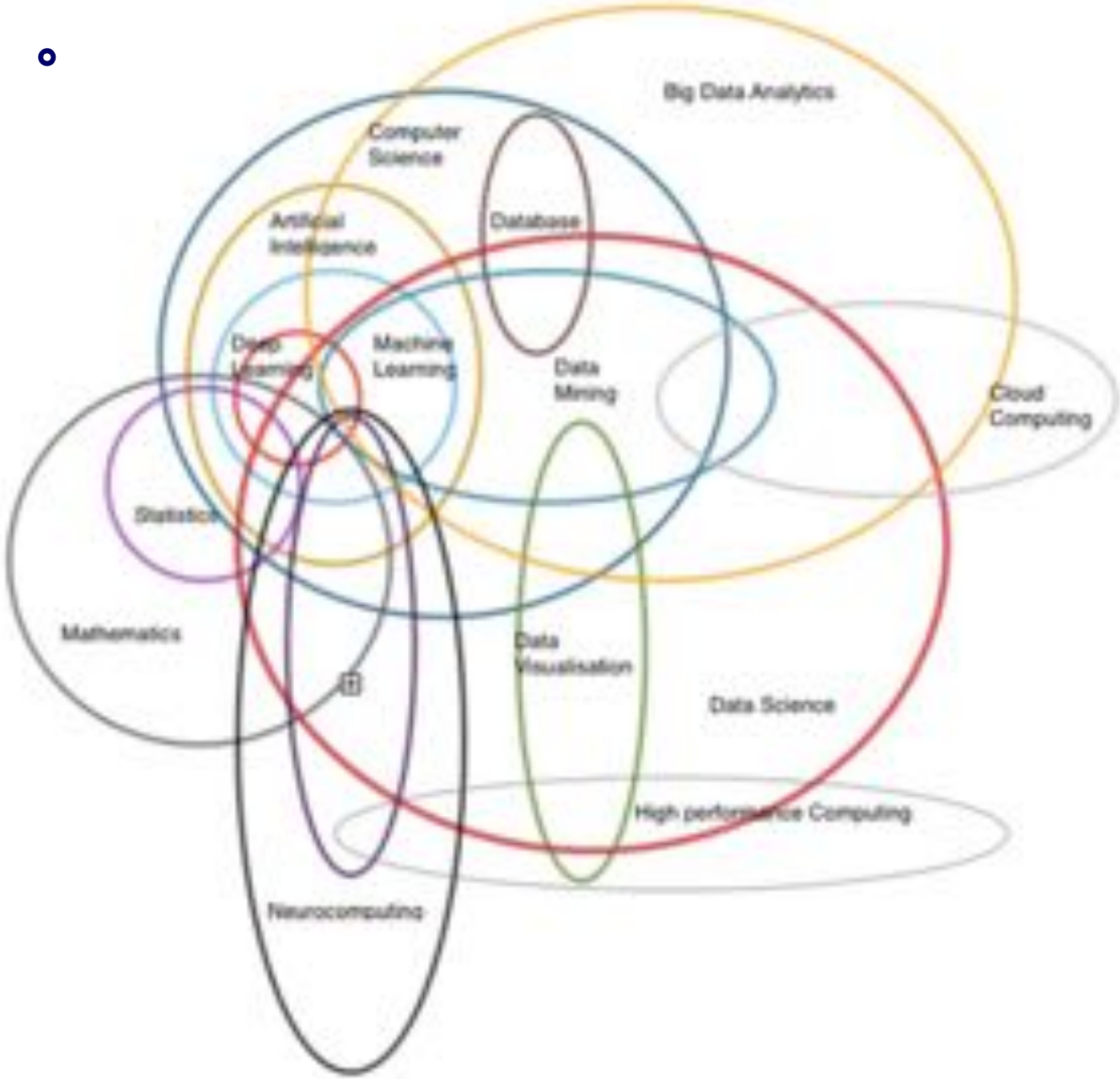

Deep learning
Example:
MLPs
Example:
Shallow
autoencoders
Example:
Logistic
regression
Example:
Knowledge
bases
Representation learning
Machine learning
AI

机器学习 与 数据挖掘和大数据

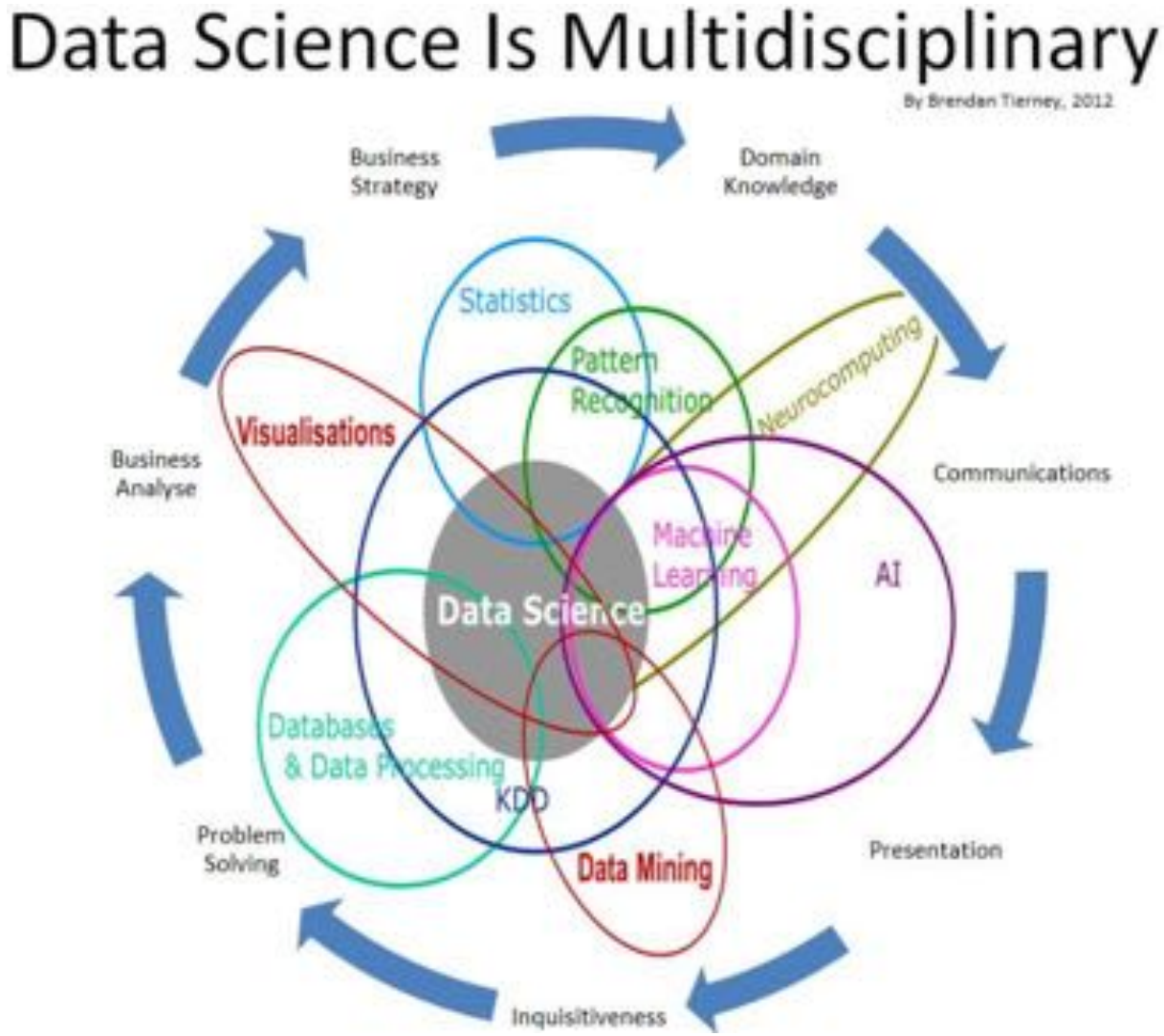


机器学习 与。。。

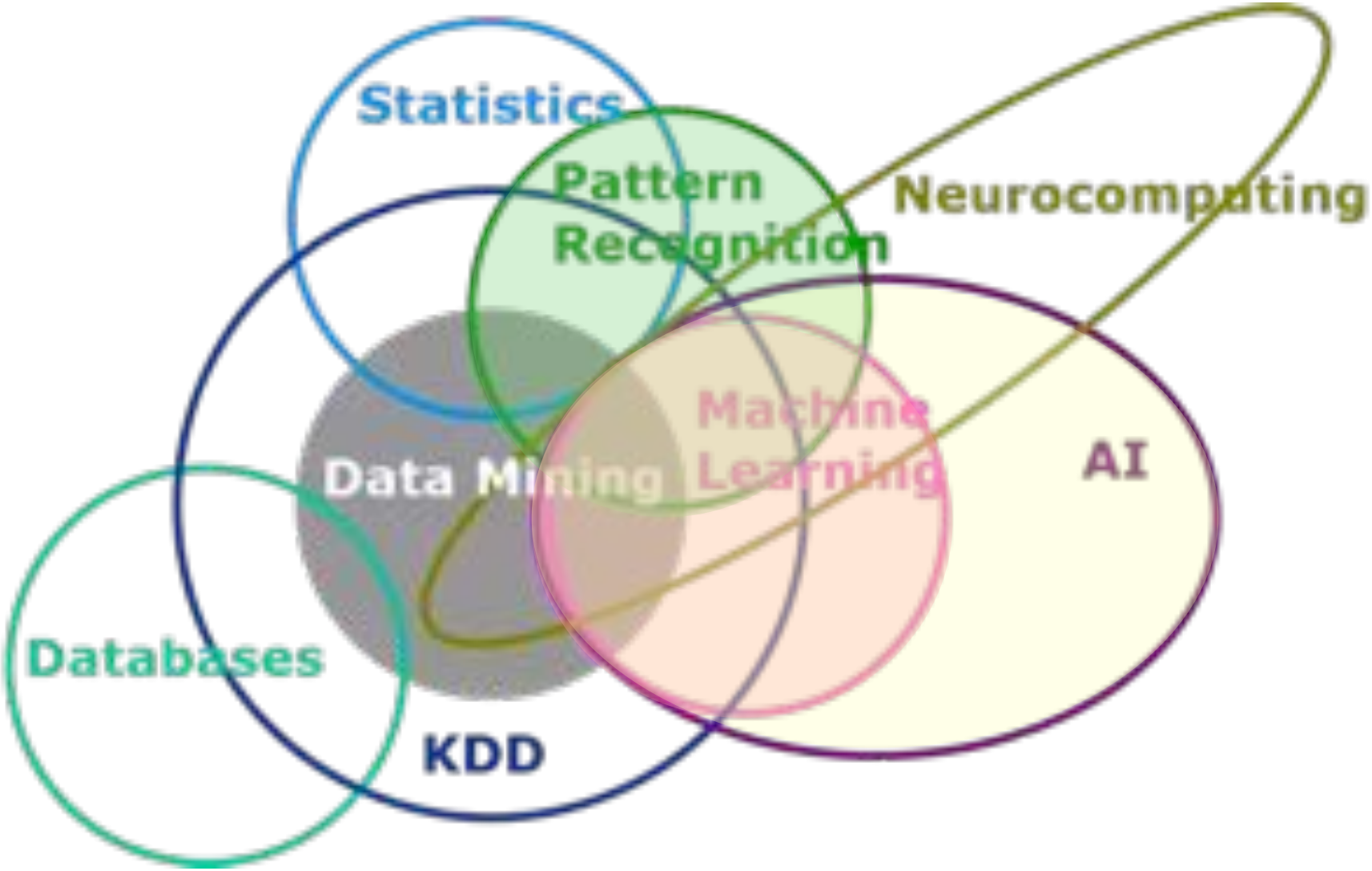
史春奇


Big Data Analytics
Computer Science
Artificial Intelligence
Database
Machine Learning
Data Mining
Cloud Computing
Statistics
Mathematics
Data Visualisation
Data Science
High performance Computing
Neurocomputing

数据科学

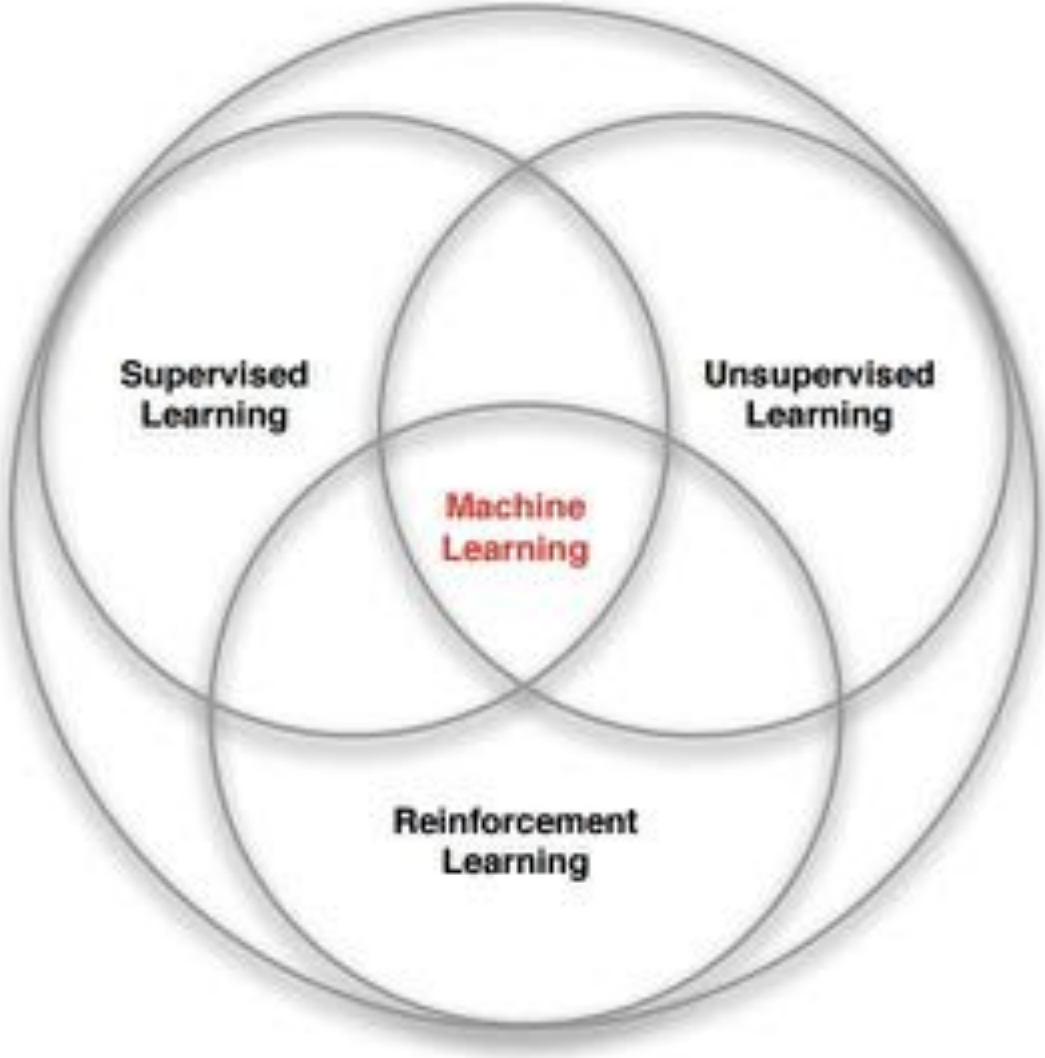

Data Science Is Multidisciplinary
By Brendan Tierney, 2012
Business Strategy
Domain Knowledge
Communications
Presentation
Inquisitiveness
Problem Solving
Business Analyse
Statistics
Pattern Recognition
Neurocomputing
Visualisations
Machine Learning
AI
Data Science
Databases & Data Processing
KDD
Data Mining

模式识别和人工智能的碰撞



机器学习的三大传统主流

- **监督学习**
- 无监督学习
- 强化学习



HOW ?

机器学习是如何改变世界的？

近二十年热点事件

谷歌Alpha GO 2.0 – 围棋

2017年5月23日到27日

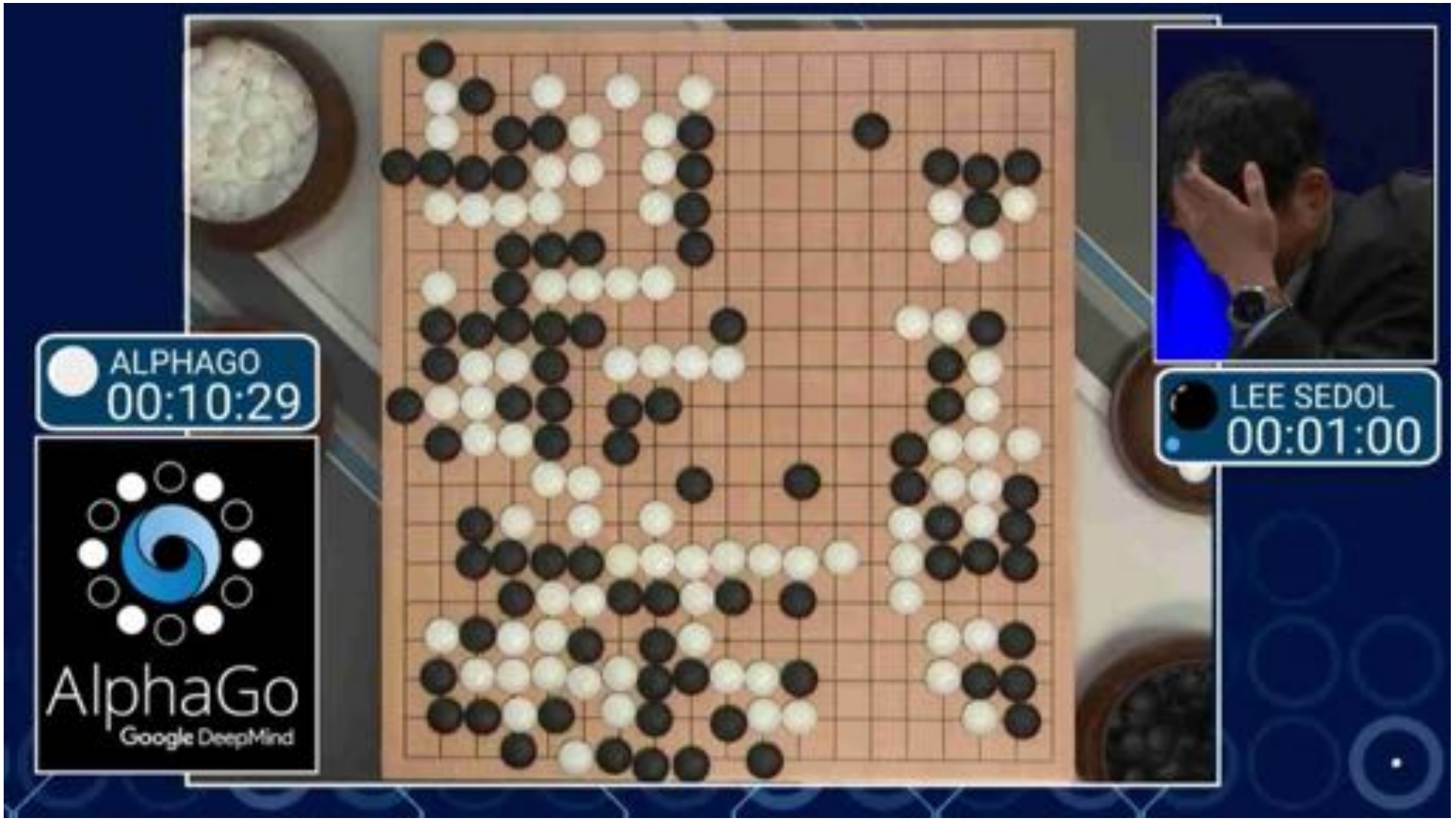
在中国乌镇围棋峰会上，阿尔法围棋以3比0的总比分战胜排名世界第一的世界围棋冠军柯洁。

CMU Libratus – Poker 德州扑克

2017年1月11日至1月30日

卡耐基梅隆大学（CMU）开发的AI程序Libratus 击败人类顶级职业玩家，赢取了20万美元的奖金。

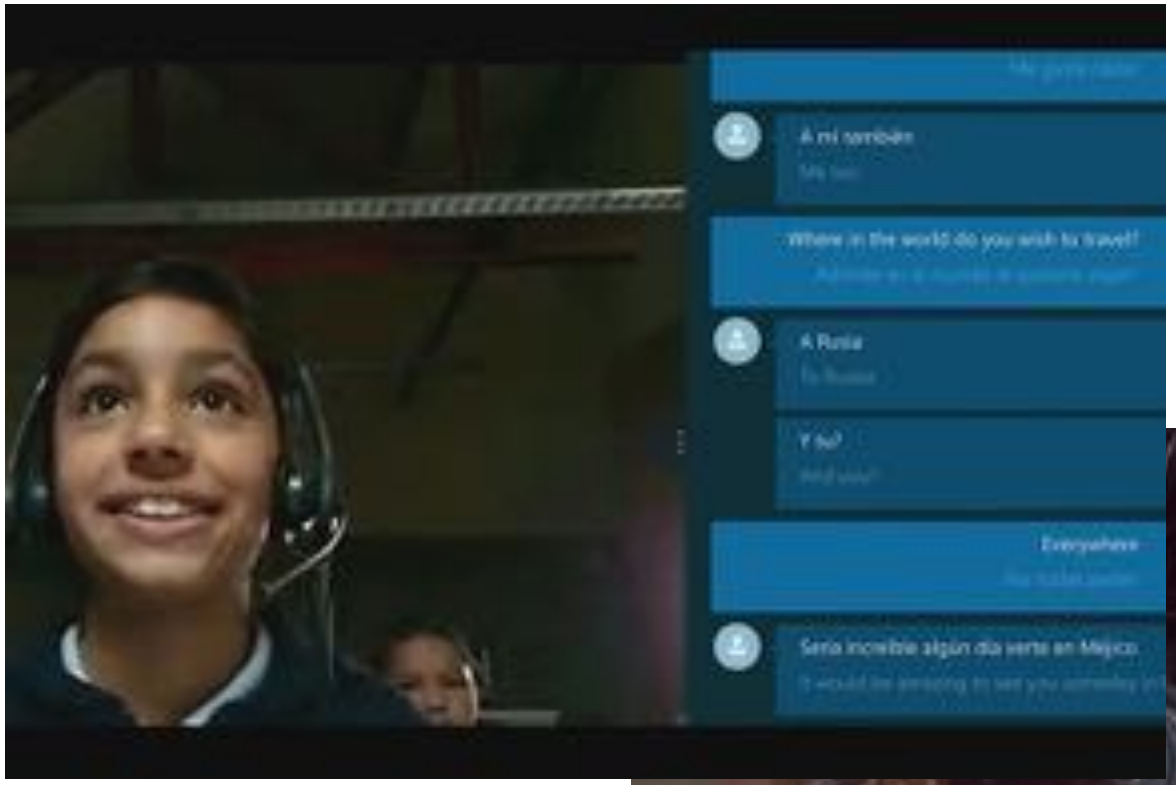
谷歌Alpha GO – 围棋

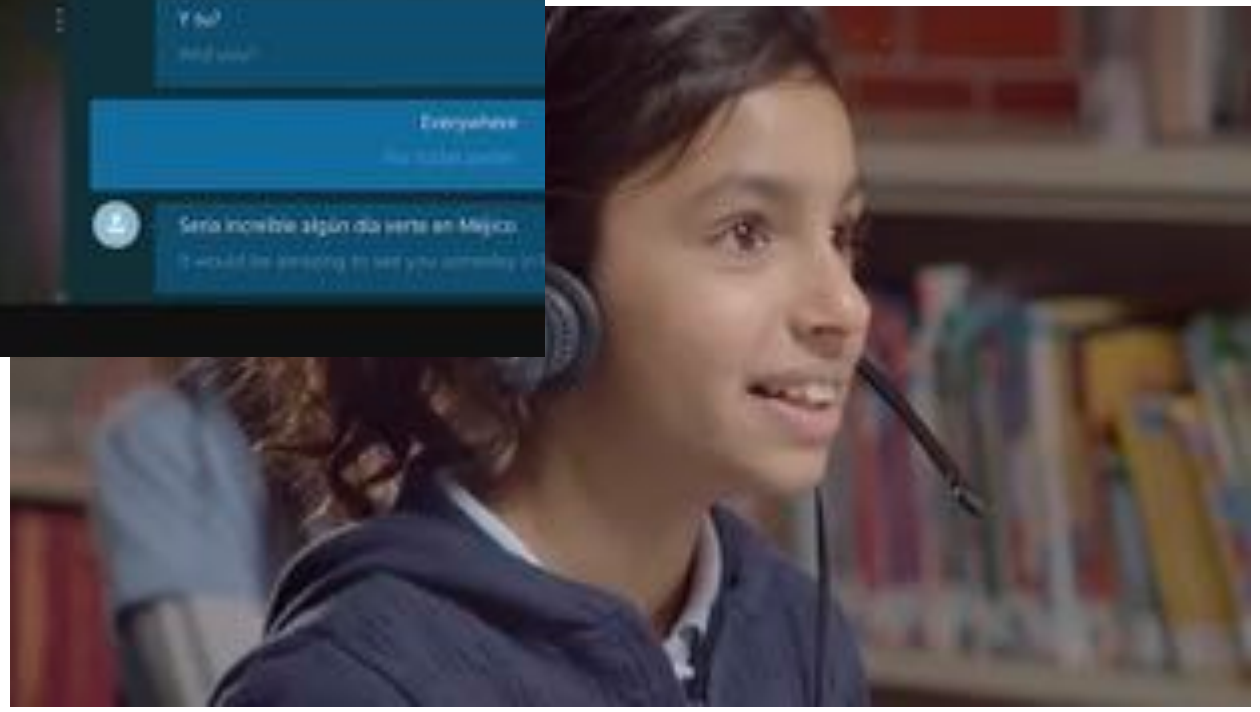


2016年3月

阿尔法围棋与围棋世界冠军、职业九段棋手李世石进行围棋人机大战，以4比1的总比分获胜；

微软 Skype – 语音翻译





2014年12月

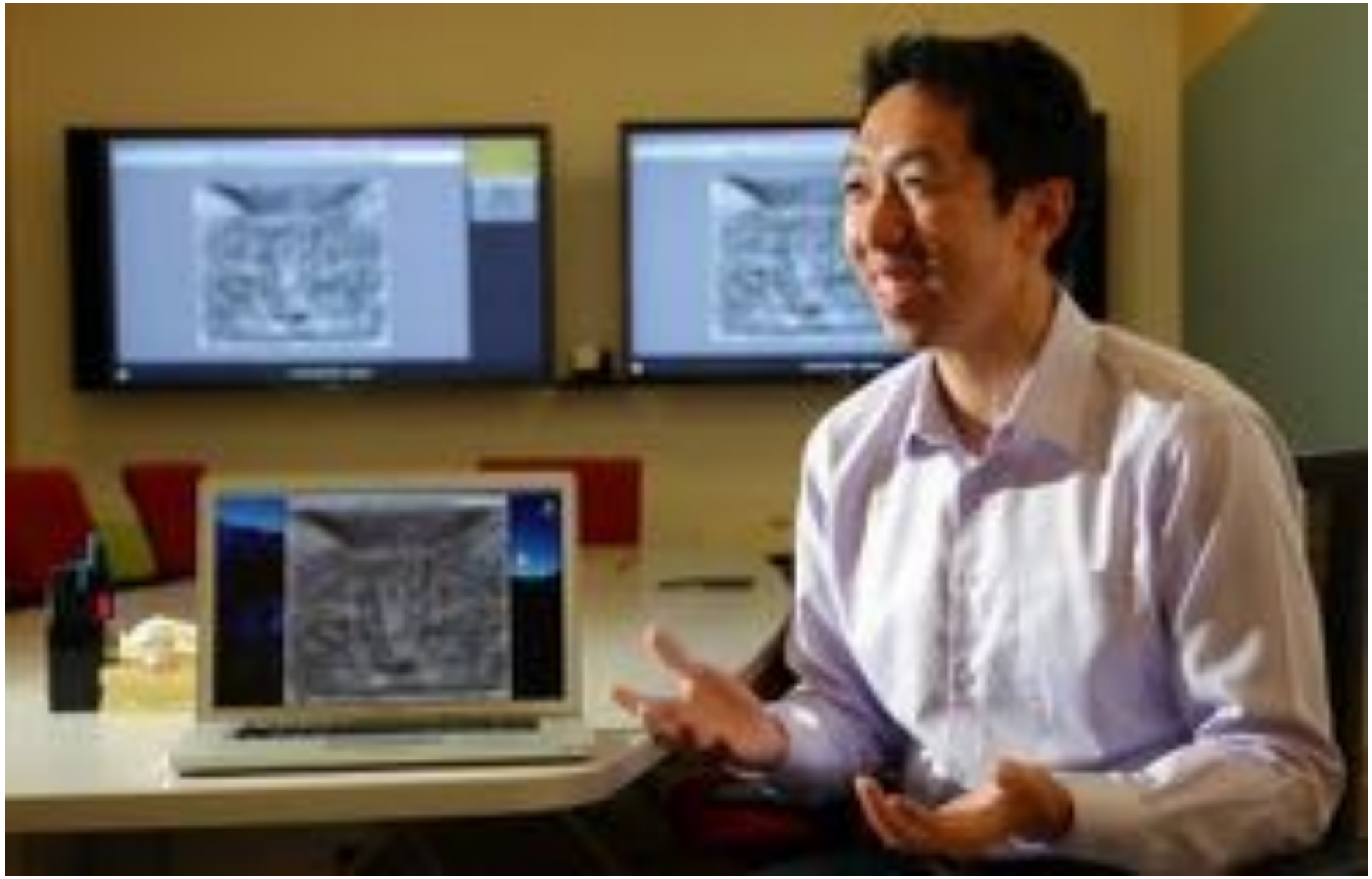
Skype上线语音到语音的机器翻译，可以实时不同语言对话。

Facebook DeepFace – 人脸识别

2014年3月

Facebook人脸识别系统DeepFace达到 97.35%准确率， 超过人的识别准确率。

谷歌大脑 – 自主物体识别



2012年6月

谷歌大脑从Youtube视频自主学习出猫。

Google X部门的科学家们通过将1.6万台电脑的处理器相连接建造出了全球为数不多的最大中枢网络系统，它能自主学习，可以称之谓“谷歌大脑”。

苹果 Siri – 人机对话

2011年10月

Siri实现了语音识别功能

IBM Watson – 电视问答智力竞赛

2011年2月14日

“危险边缘”Jeopardy!是哥伦比亚广播公司一档长盛不衰的电视问答节目。Watson赢得100万美元奖金而一举成名。

这些得以实现靠的是90台IBM服务器、360个计算机芯片驱动以及IBM研发的DeepQA系统。

谷歌 Waymo – 自动驾驶车

2009年

谷歌开启自动驾驶车的项目

2012年5月8日，在美国内华达州允许无人驾驶汽车上路3个月后，机动车管理局为Google的无人驾驶汽车颁发了一张合法车牌。为了醒目的目的，无人驾驶汽车的车牌用的是红色。

Netflix大奖赛 – DVD自动推荐

2006年 – 2009年

美国DVD租赁公司启动100万美金的Netflix大奖赛 。

Bellkor小组融合了数百种预测模型被融合到一起，最终突破了0.8567的RMSE的精度。

全球186国家的41,305个组队提交了 51,051个算法。

史春奇

谷歌 – 机器翻译

2006年

谷歌机器翻译上线。

到目前谷歌翻译的准确率大幅度提高，并且支持图片上的文字翻译。

iRobot Roomba – 扫地机器人

2002年

全球第一款家用清洁机器人 iRobot Roomba 正式上市。

到2011年就卖了600万台 Roomba。

IBM 深蓝 – 国际象棋

1997年

IBM 的超级计算机深蓝（Deep Blue）和有史以来最神奇的国际象棋世界冠军卡斯帕罗夫展开了六盘人机大战。

人机大战六盘，深蓝最终以 3.5 比 2.5 胜出，这是人类历史上计算机第一次在国际象棋六番棋中战胜人类的世界冠军。

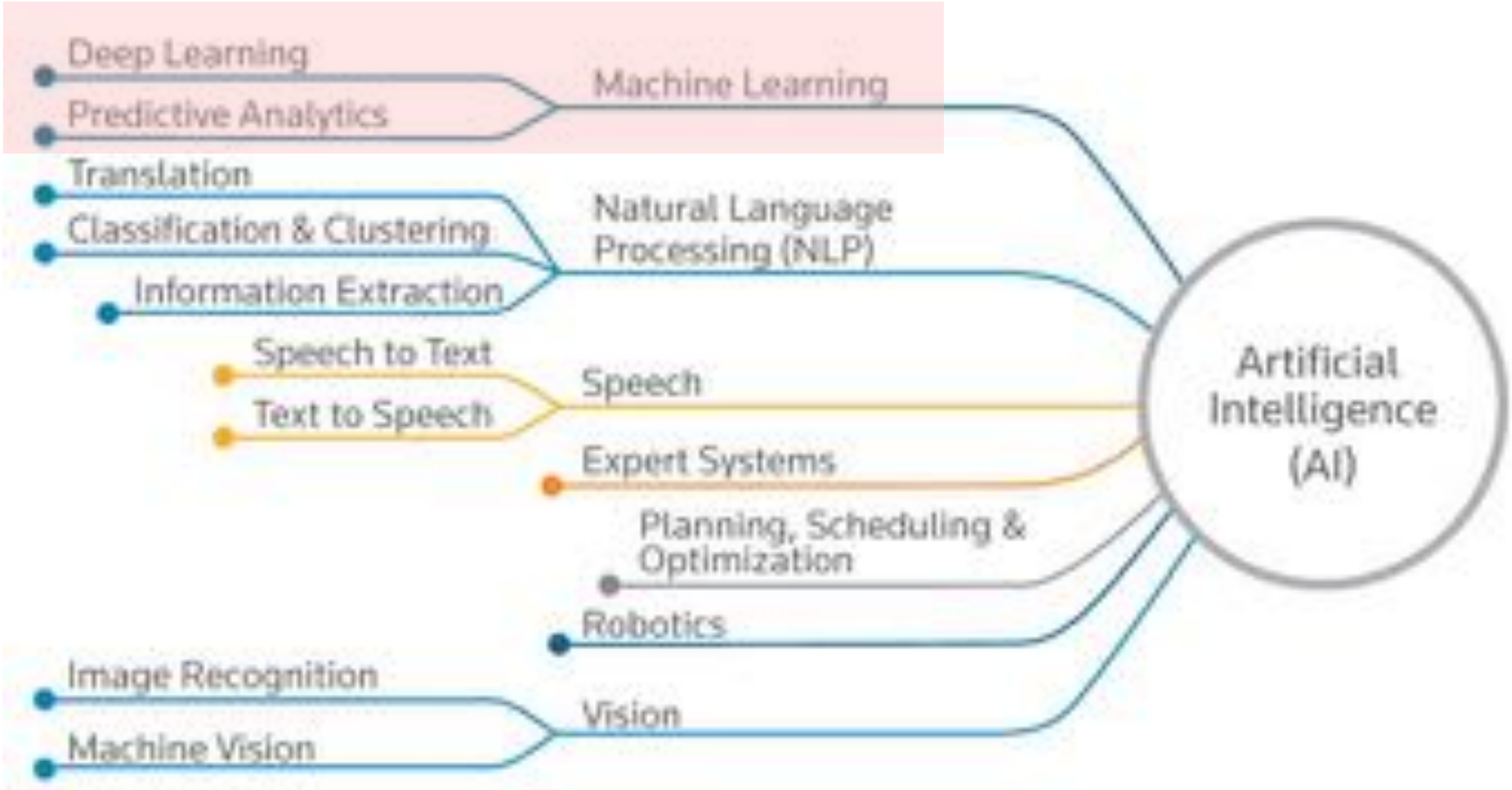
WHEN

- 什么时候机器开始发力？

人工智能由来

人工智能（Artificial Intelligence）

- Artificial intelligence (AI) is intelligence exhibited by machines.

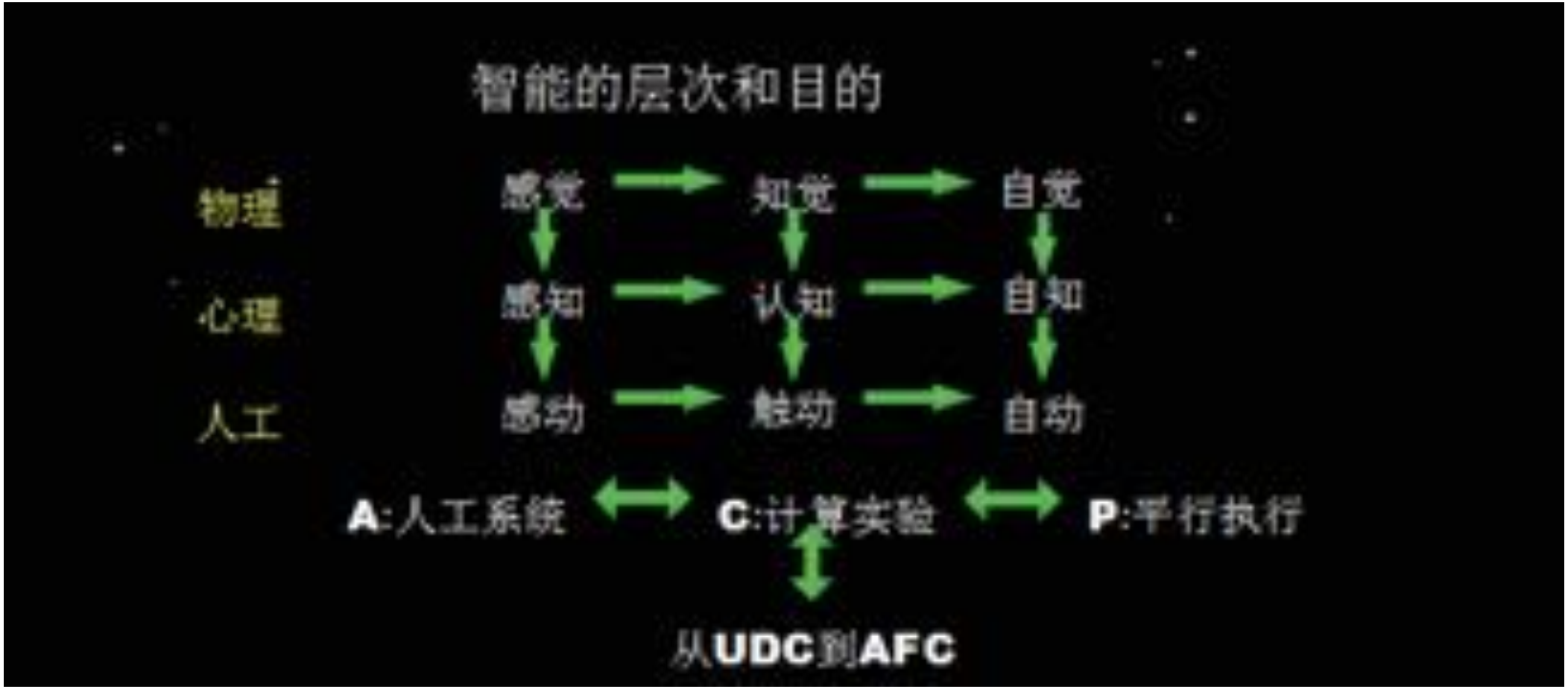


WHAT

- 什么是机器智能？

chunqi.shi@hotmail.com chunqishi
史春奇

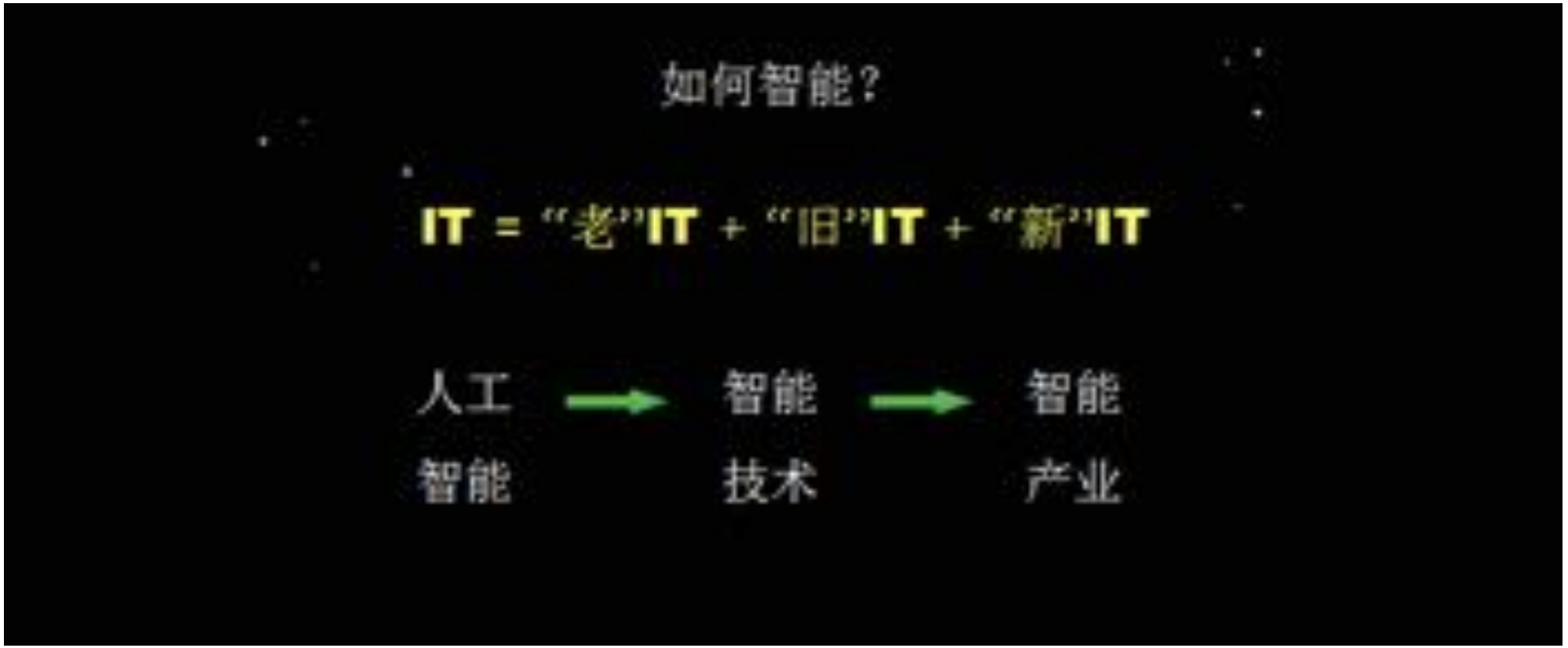
智能是什么？



王飞跃

chunqi.shi@hotmail.com chunqishi
史春奇

智能是什么？

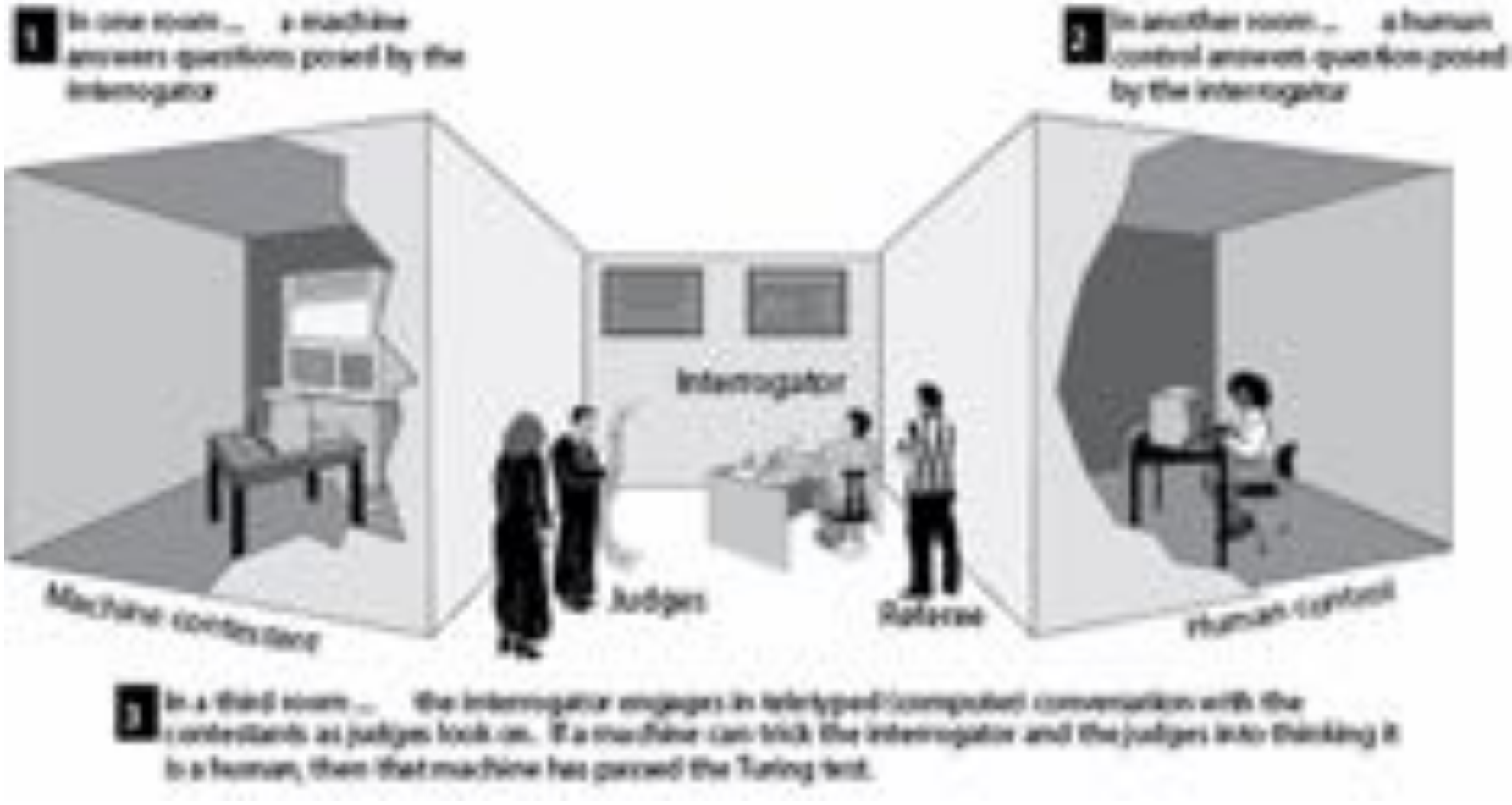


chunqi.shi@hotmail.com chunqishi
史春奇

智能世界

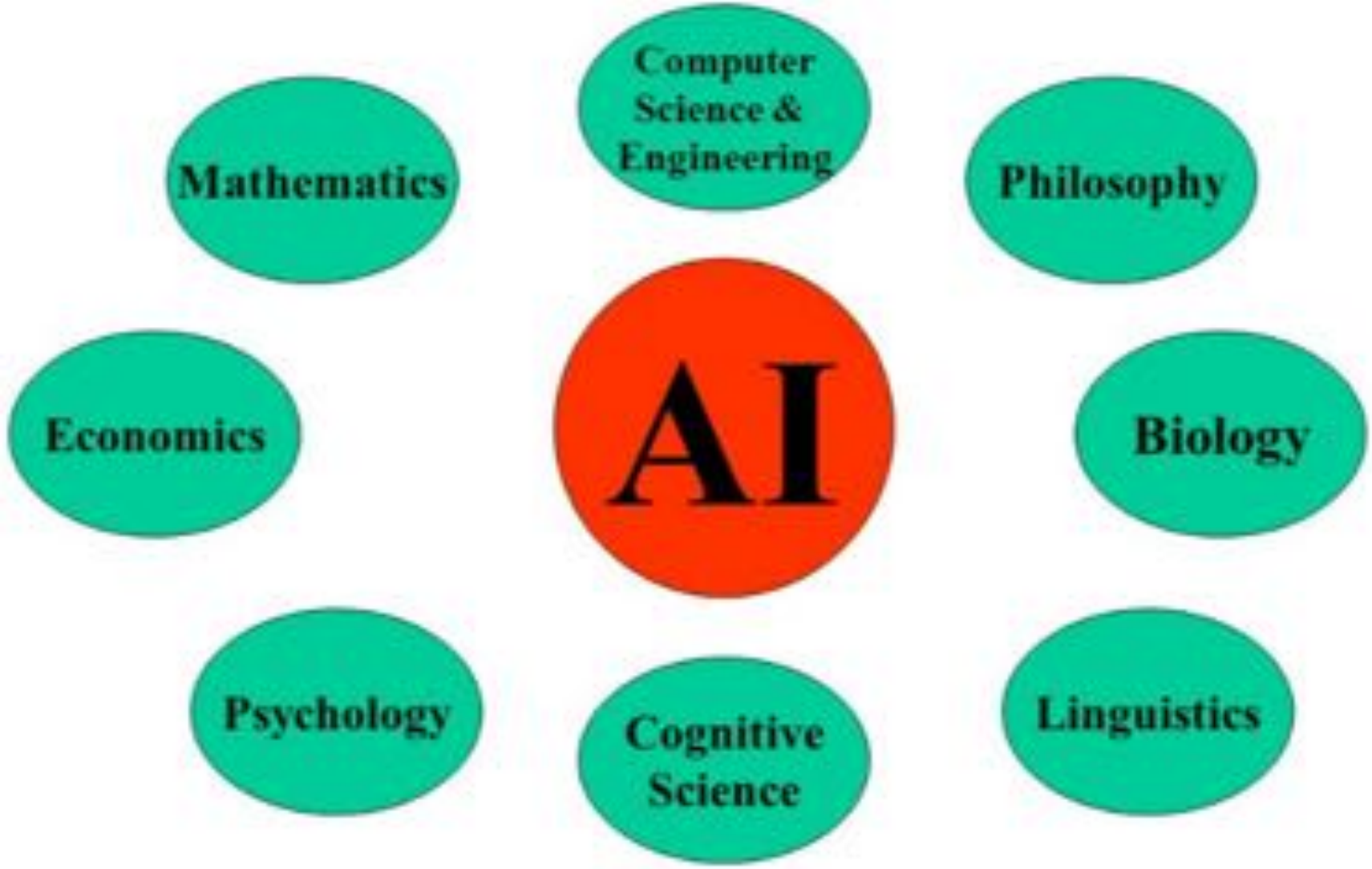
是否有智能？

- 图灵测试



chunqi.shi@hotmail.com chunqishi
史春奇

人工智能（Artificial Intelligence）



哲学 (400 B.C-) – 起源

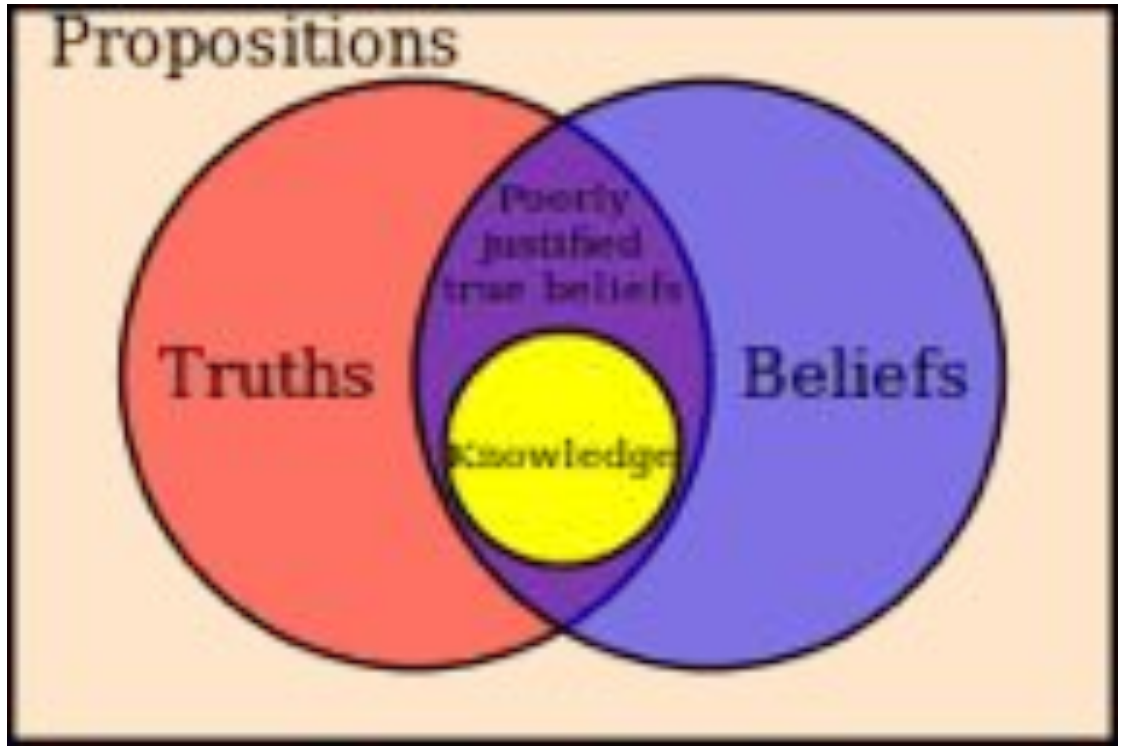
- Socrates → Plato → Aristotle

Standard whereby to judge your actions and those of other men" (**algorithm**)

Formulate laws of rational part of the mind

哲学: 知识来源 – 经验主义

- Empiricism (Francis Bacon 1561-1626)



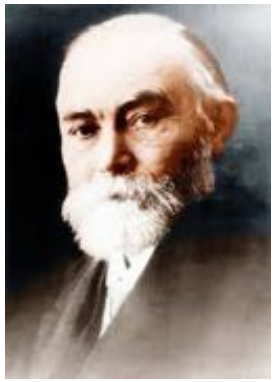
哲学: 知识来源 – 实证主义

- Logical Positivism（ Bertrand Russell 1872-1970 ）

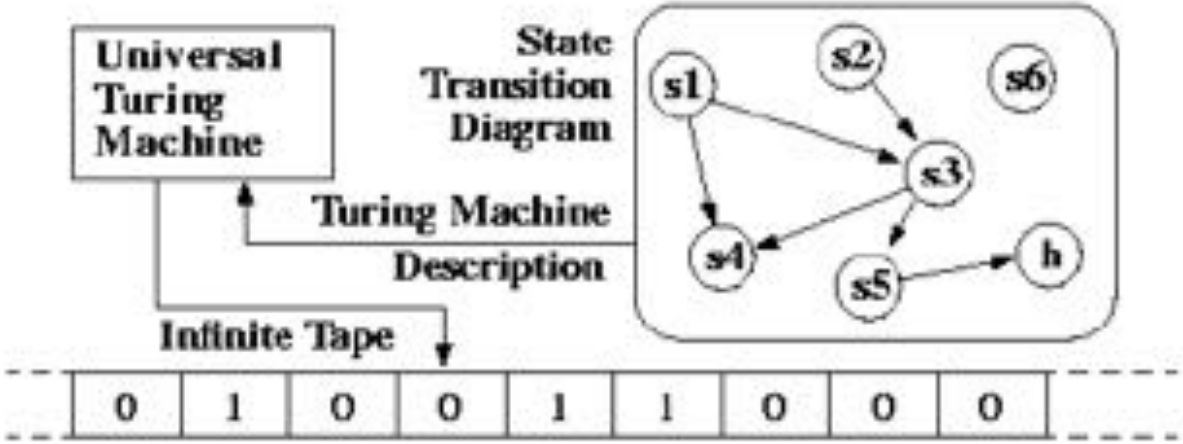
	THOUGHT	OBJECTS	
KNOWLEDGE	REASON (DIALECTIC)	HIGHER FORMS	INTELLIGIBLE WORLD
	UNDERSTANDING SCIENCE AND MATHEMATICS	FORMS OF SCIENCE AND MATHEMATICS	
OPINION	BELIEF (PERCEPTION)	THINGS, OBJECTS	VISIBLE WORLD
	CONJECTURE (IMAGINATION)	SHADOWS, IMAGES, REFLECTIONS	

数学

- 逻辑
  - George Boole (1815-1864)
    - Formal language
  - Gottlob Frege (1848-1925)
    - First-order logic

- 可计算性
  - David Hilbert (1862-1943)
    - Problem #23
  - Alan Turing (1912-1954)
    - Turing machine



数学

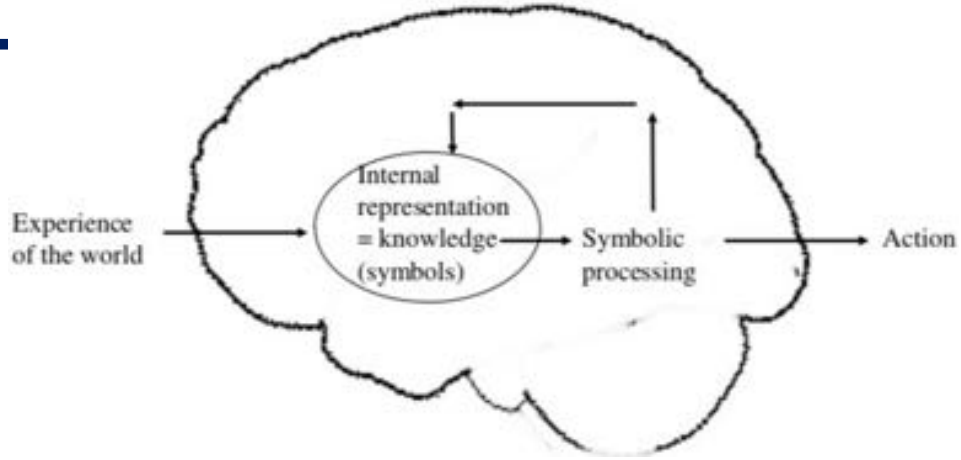
- Probability
  - Thomas Bayes (1702-1761)

- Decision theory
  - John Von Neumann（1903-1957）

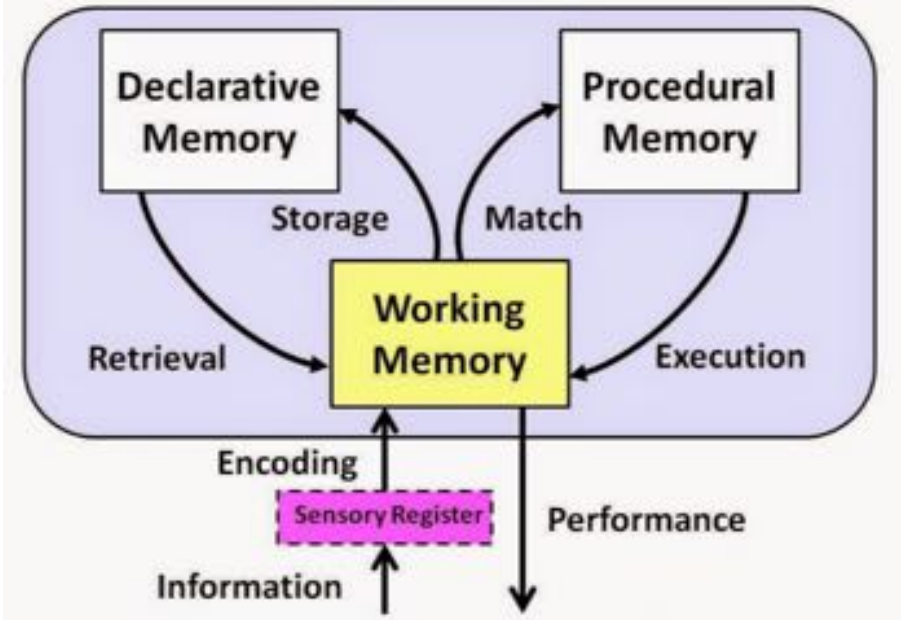
- Game theory
  - Jean Tirole（1953-）
    - Nobel Prize in Economic Sciences in 2014

心理学

- Cognitive psychology
  - Kenneth Craik (1914–1945)
    - knowledge-based agent



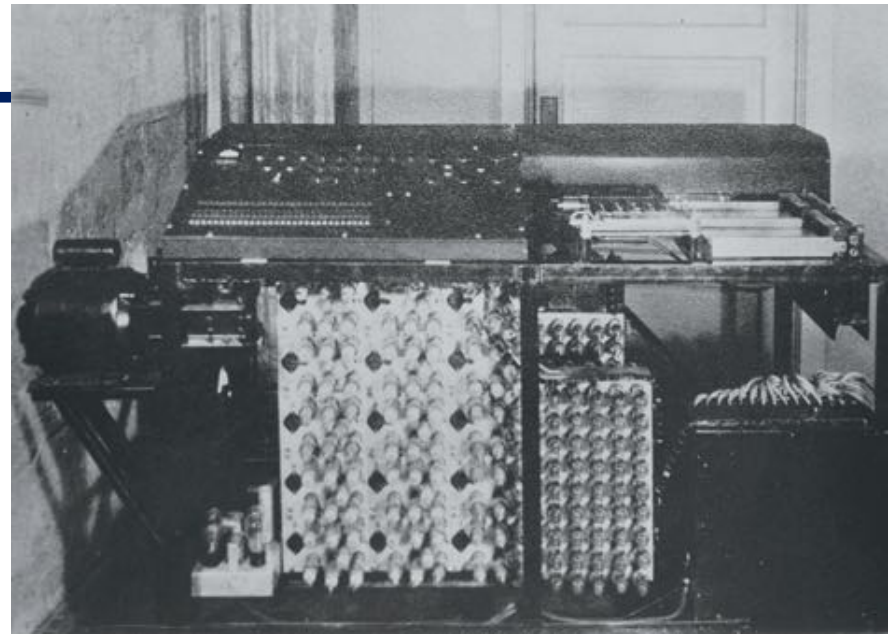
  - John B. Anderson （1922- ）
    - Adaptive concept of thought



chunqi.shi@hotmail.com chunqishi
史春奇

计算机

- ABC （1942）

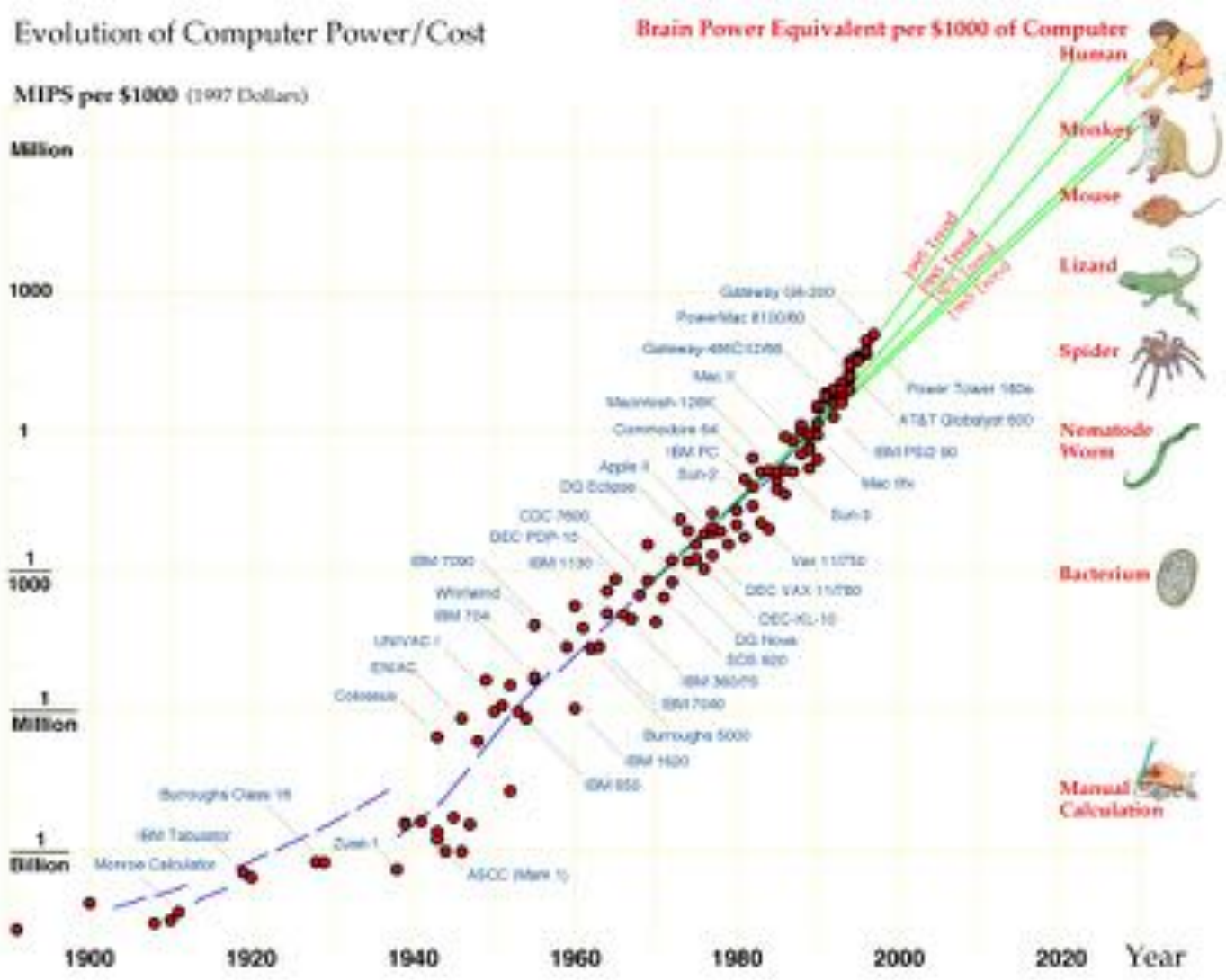


- ENIAC （1946）

chunqi.shi@hotmail.com chunqishi

史春奇

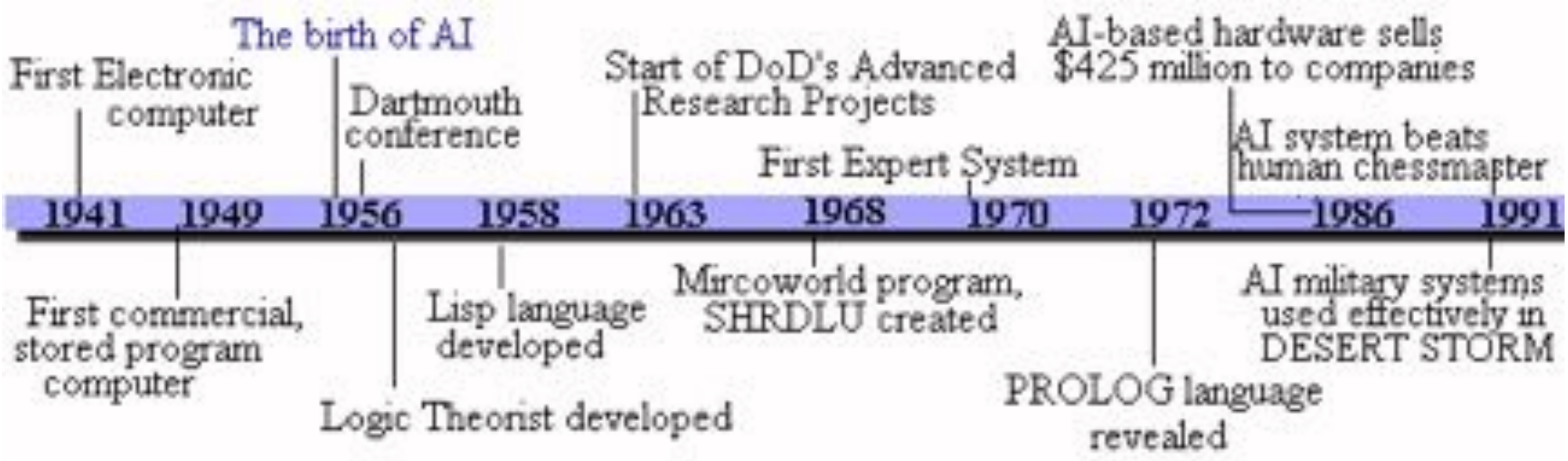
计算机能力的演变



chunqi.shi@hotmail.com chunqishi

史春奇

人工智能诞生



Dartmouth Conferences （1956）

- The field of AI research was founded
  - John McCarthy, Marvin Minsky, Nathaniel Rochester and Claude Shannon

Photographer: Joe Mehling

Figure 1. Trenchard More, John McCarthy, Marvin Minsky, Oliver Selfridge, and Ray Solomonoff.

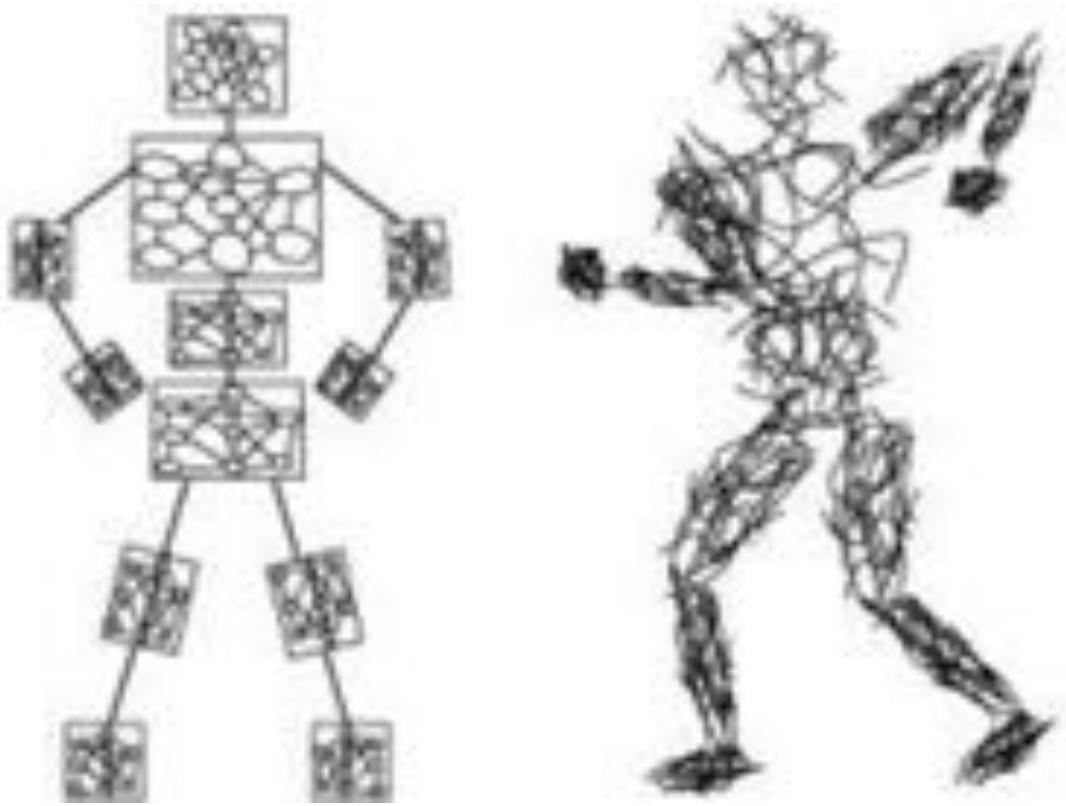
人工智能的大阶段

- 起步的狂热 (1952-1969)
  - 感知机 Perceptron (Frank Rosenblatt)
- 现实的刹车 (1966-1974)
  - XOR 问题（ 1-layer ANNs Marvin Minsky ）
- 知识的系统 (1969-1979)
  - DENDRAL 和 Mycin
- 产业的萌生 (1980-1988)
  - 日本五代机项目
- 神经网络的回归 (1986-)
  - Backpropagation

chunqi.shi@hotmail.com chunqishi
史春奇

人工智能方法论

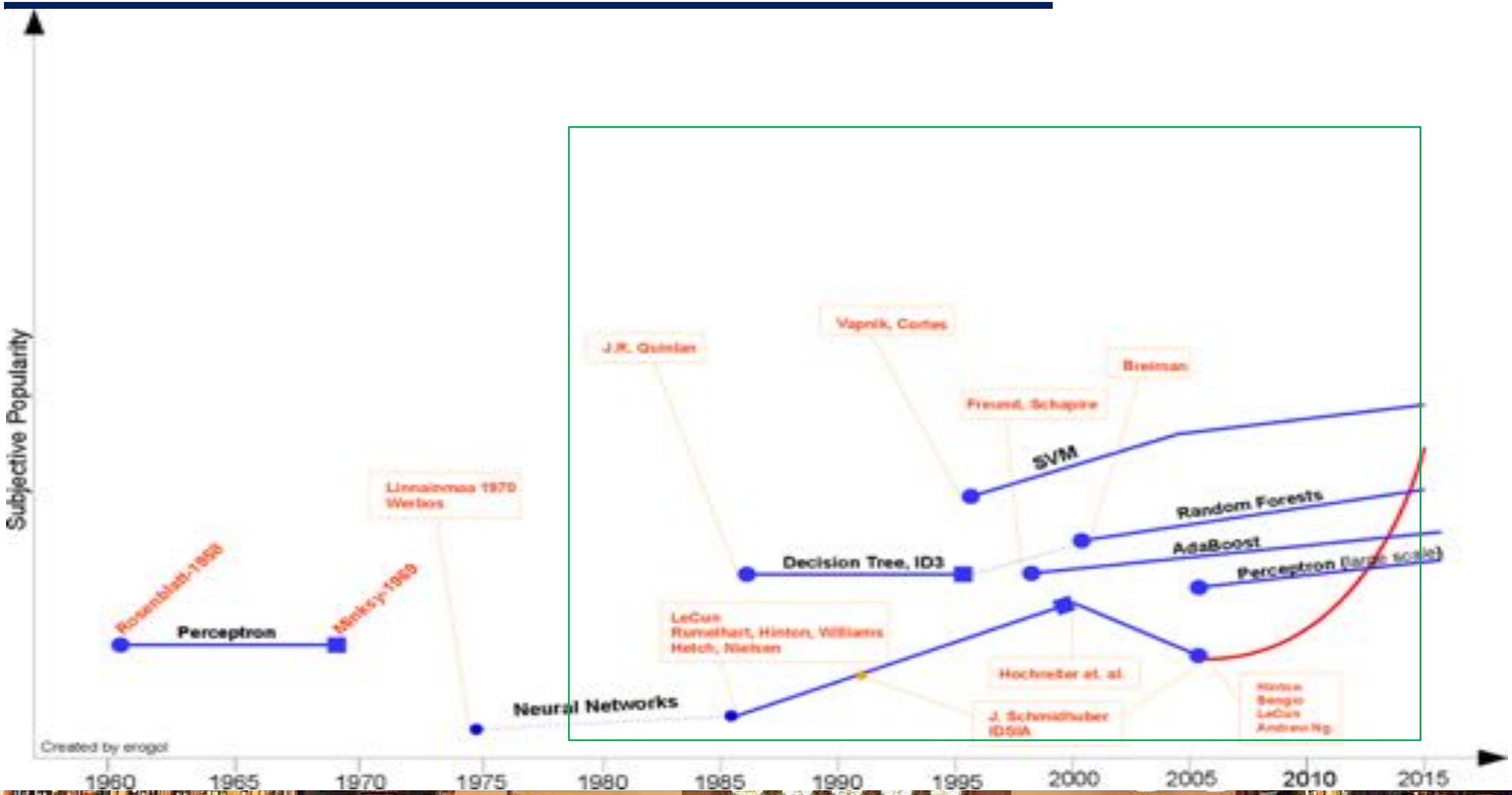
- 符号主义 (Symbolism)
- 链接主义 (Connectionist)
- 嵌入主义 (Situated)
- 进化主义 (Evolutionary)



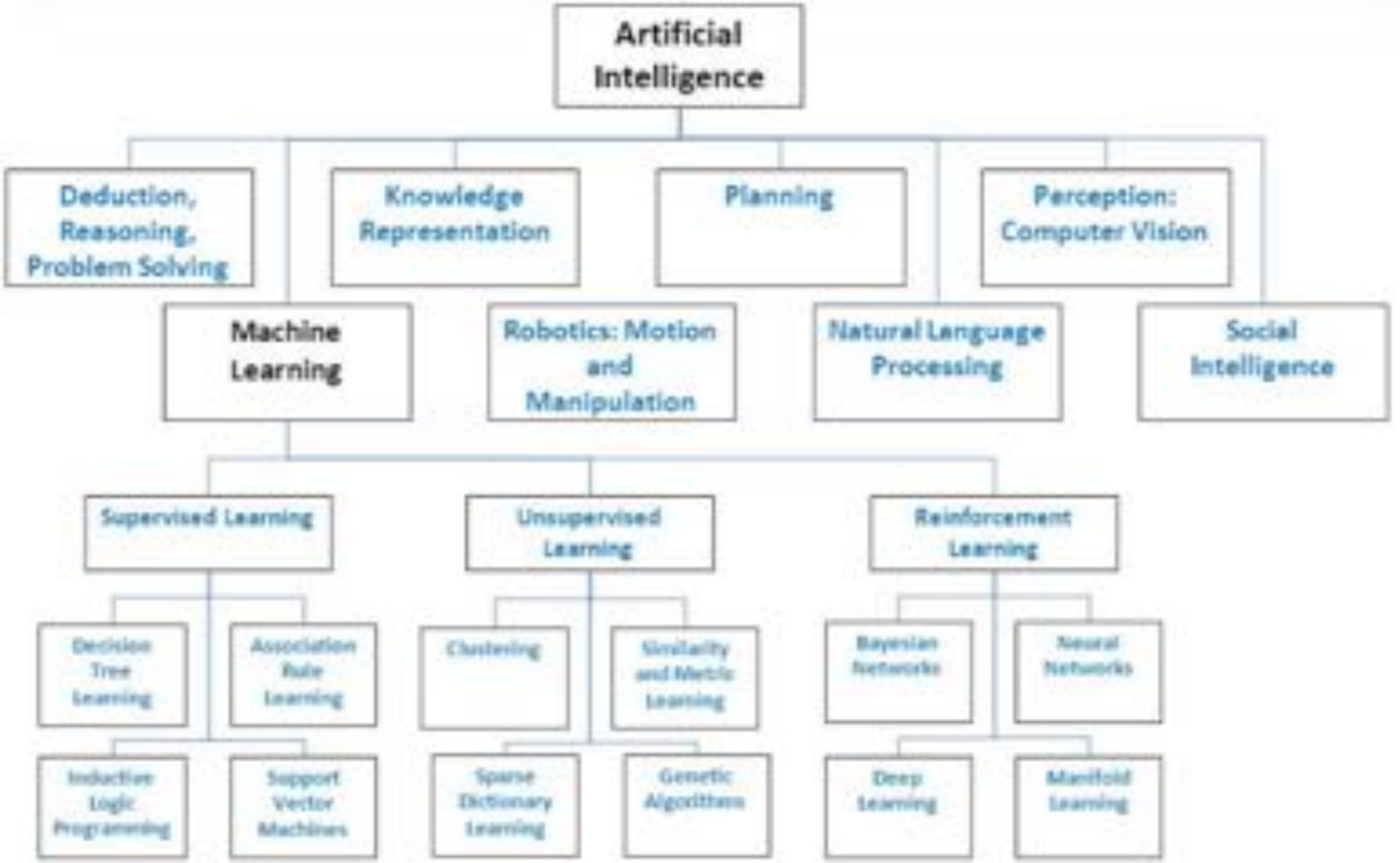
chunqi.shi@hotmail.com chunqishi

史春奇

统计学习和深度学习的崛起（1986- ）



chunqi.shi@hotmail.com chunqishi
史春奇



WHY MACHINE LEARNING?

- 为什么机器学习会崛起？

可能近似正确PAC模型

- PAC Model
  - Probably Approximately Correct
  - **1984年提出**
- Leslie Valiant
  - 哈佛Harvard大学教授
  - 受到计算复杂性理论（Computational complexity theory）启发
  - 2010年图灵奖获得者

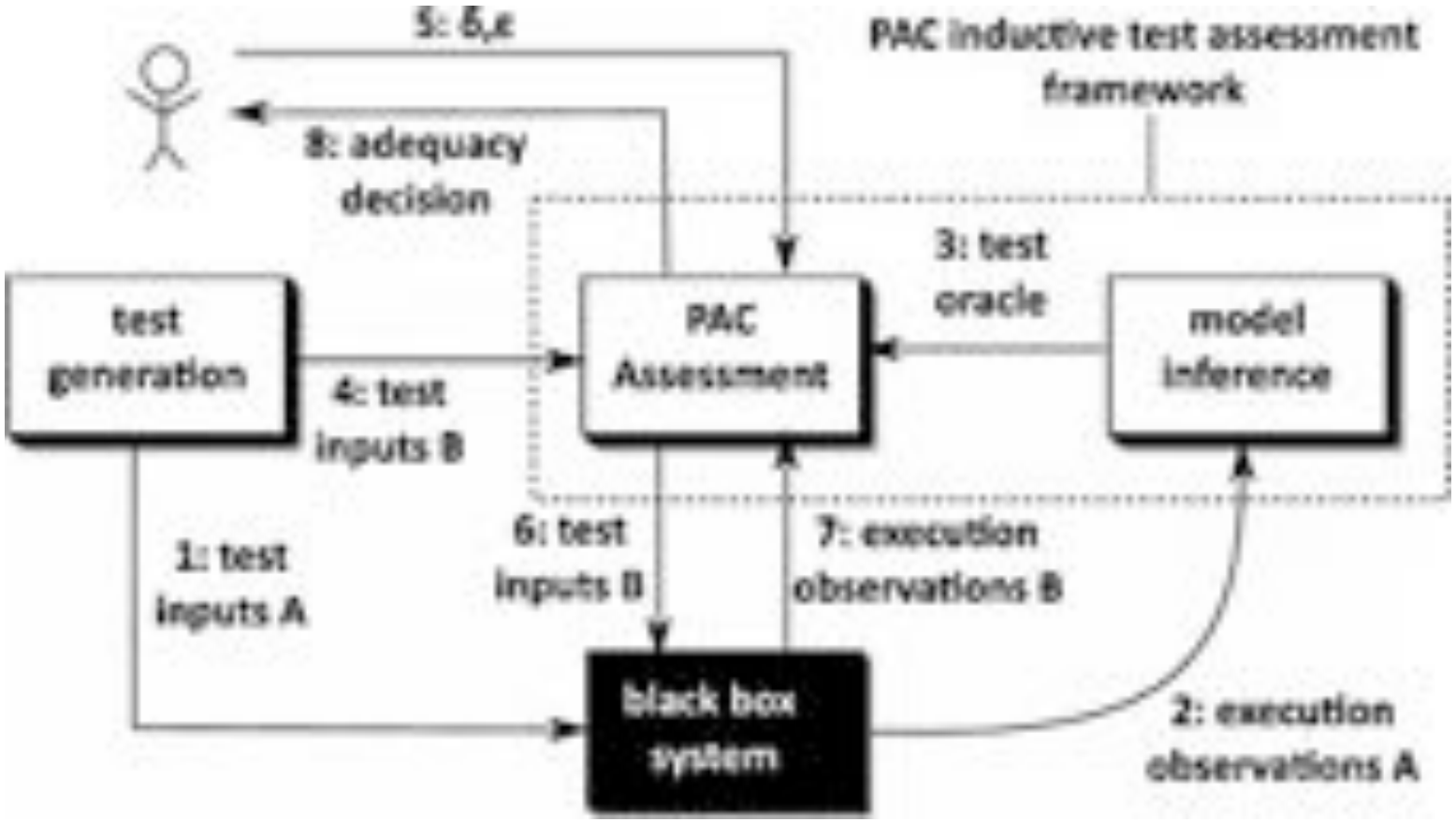
WHY PAC?

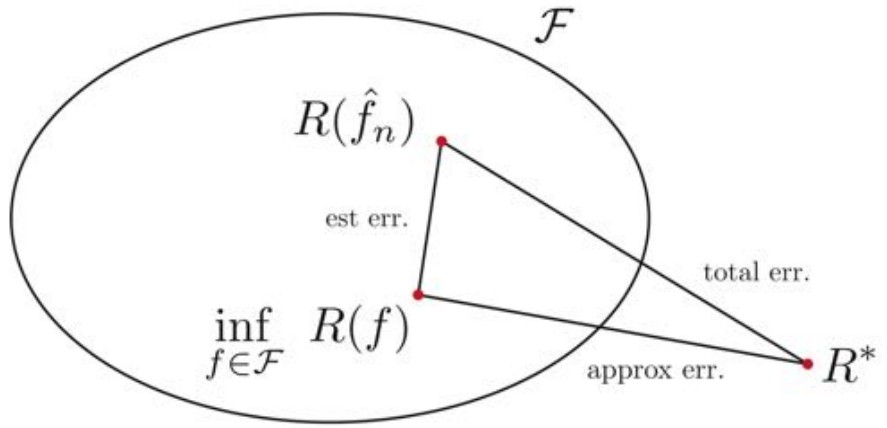
- PAC是机器学习理论基础
  - 基本概念都是基于PAC泛化理论的。
  - 包括： 损失函数、风险函数、经验结构最小化、结构风险最小化、学习方法的泛化能力、VC维

- PAC获得图灵奖的机器学习内容

2010	Leslie G. Valiant	For transformative contributions to the theory of computation, including the theory of probably approximately correct (PAC) learning, the complexity of enumeration and of algebraic computation, and the theory of parallel and distributed computing.

WHAT PAC?





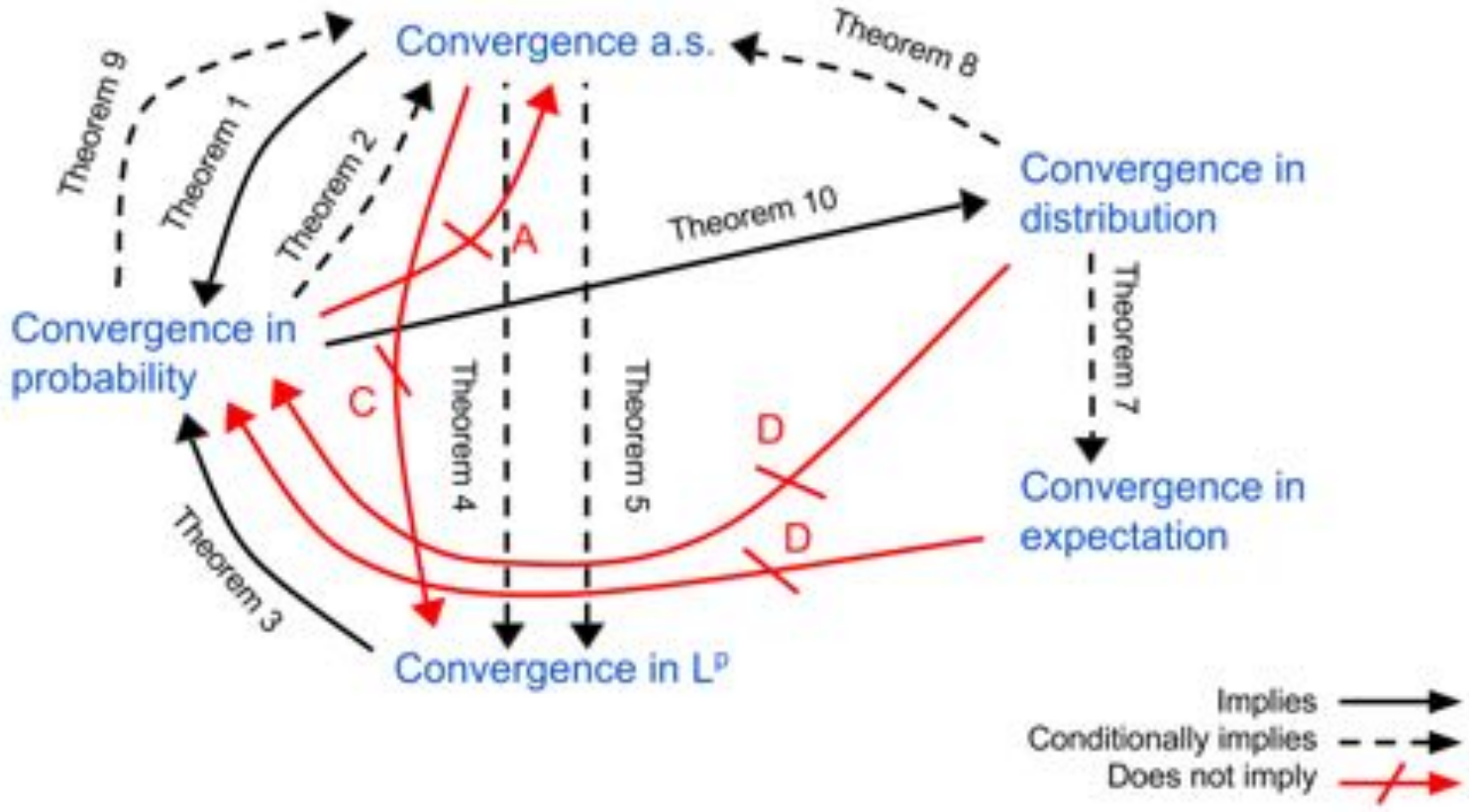
$$\Pr[error_D(h) < \varepsilon] > 1 - \delta$$

PAC可学习的基础

- 概率统计的发展！

$$\Pr[error_D(h) < \varepsilon] > 1 - \delta$$

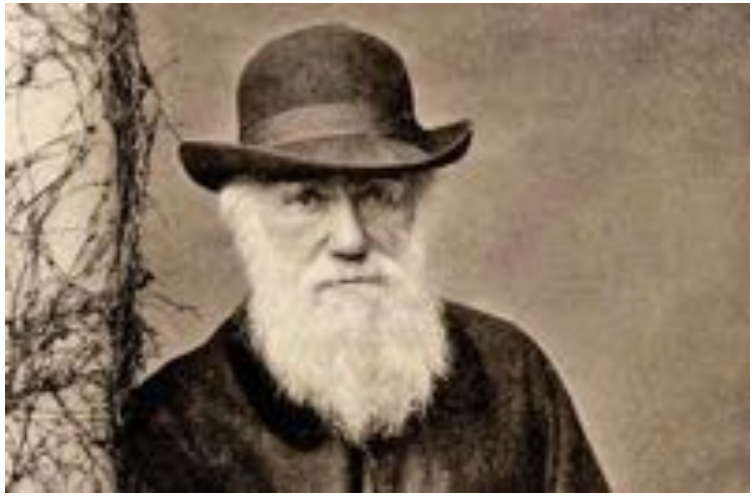
- 统计工具的使用！



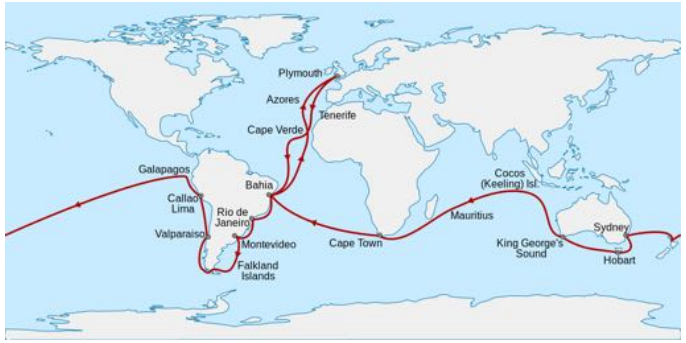
统计由来

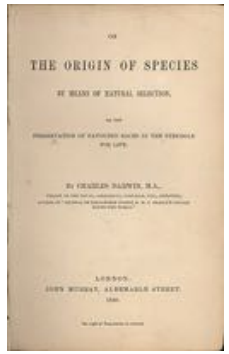
表兄弟

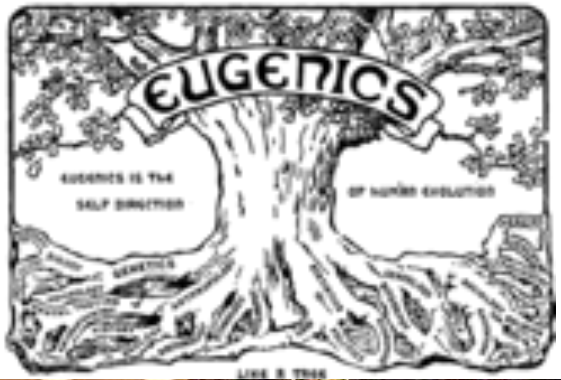
达尔文 Darwin



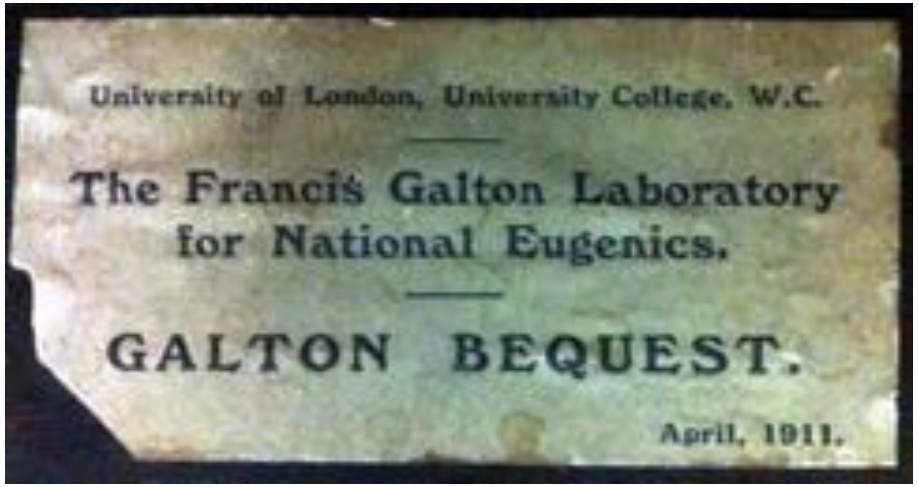
高尔顿 Galton




THE ORIGIN OF SPECIES



高尔顿优生实验室



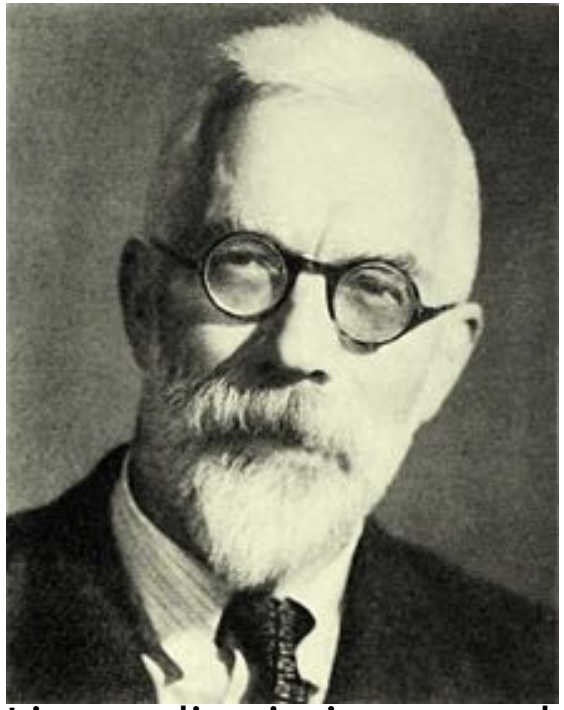
- Karl Pearson 1911-1933
- Ronald Fisher 1933-1943
- Lionel Penrose ?-1965
- Harry Harris 1965-1976
- Bette Robson 1976-1994
- Nick Wood (2009-present)

黄金组合

皮尔逊 Pearson

Correlation coefficient
Method of moments
Pearson's chi-squared test
Hypothesis testing theory
Principal component analysis
P value

费希尔 Fisher



Linear discriminant analysis
Sufficient statistic
Fisher information
F-distribution
Maximum likelihood estimation
Inverse probability
Analysis of variance (ANOVA)

戈塞 Gosset

t-distribution

从英国到美国

拉奥 Radhakrishna Rao

Cramer-Rao Lower Bound CRLB
Rao-Blackwell theorem

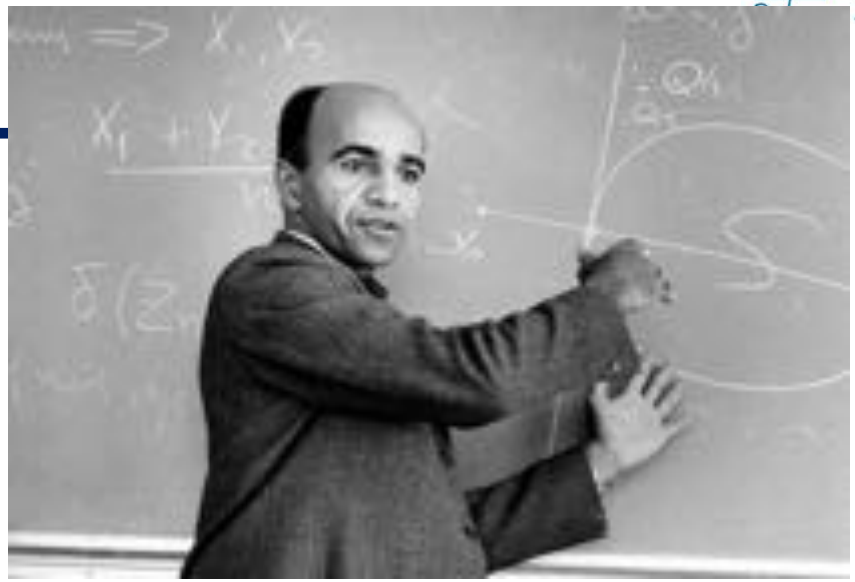
内曼 Jerzy Neyman (1894-1981)

Confidence interval
Stratified sampling
Neyman–Pearson lemma

伯克利统计系创系

- 1938年， 44 岁内曼Neyman从University College London应用统计系来到, Berkeley数学系， 开创统计实验室 Statistical Laboratory (Stat Lab)
- 1945年， 组织了 mathematical statistics and probability论坛
- 1950年， 组织了第二次论坛
- 1951年， 四位教师
  - Lehmann in 1946, Barankin in 1947, Fix and Scott in 1950 and Joseph L. Hodges in 1951
- 1954年， David Blackwell 访问伯克利统计实验室
- 1955年， 10位教师创系， Neyman是系主任
- 1956年， Blackwell是系主任

布莱克维尔 David Blackwell



- 1919年 - 2010年，美国科学院首位黑人院士、伯克利首位黑人终身教授、杰出统计学家

- University of Illinois概率博士
- 冯诺依曼邀请他讨论他的博士论文

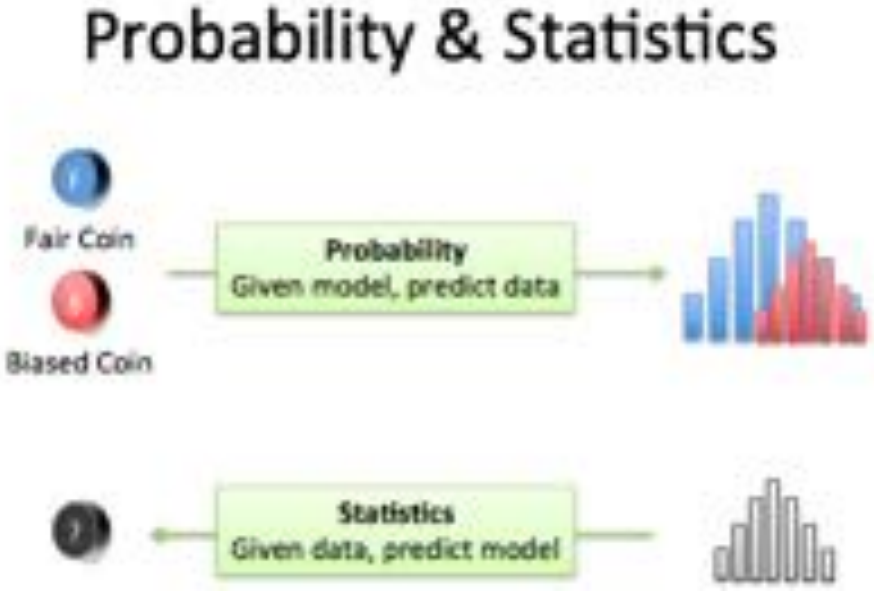
- 内曼Neyman想在伯克利数学系招聘他，未果。
- 1954年访问伯克利， 1955年统计系创系，1956年系主任, 1988年退休。

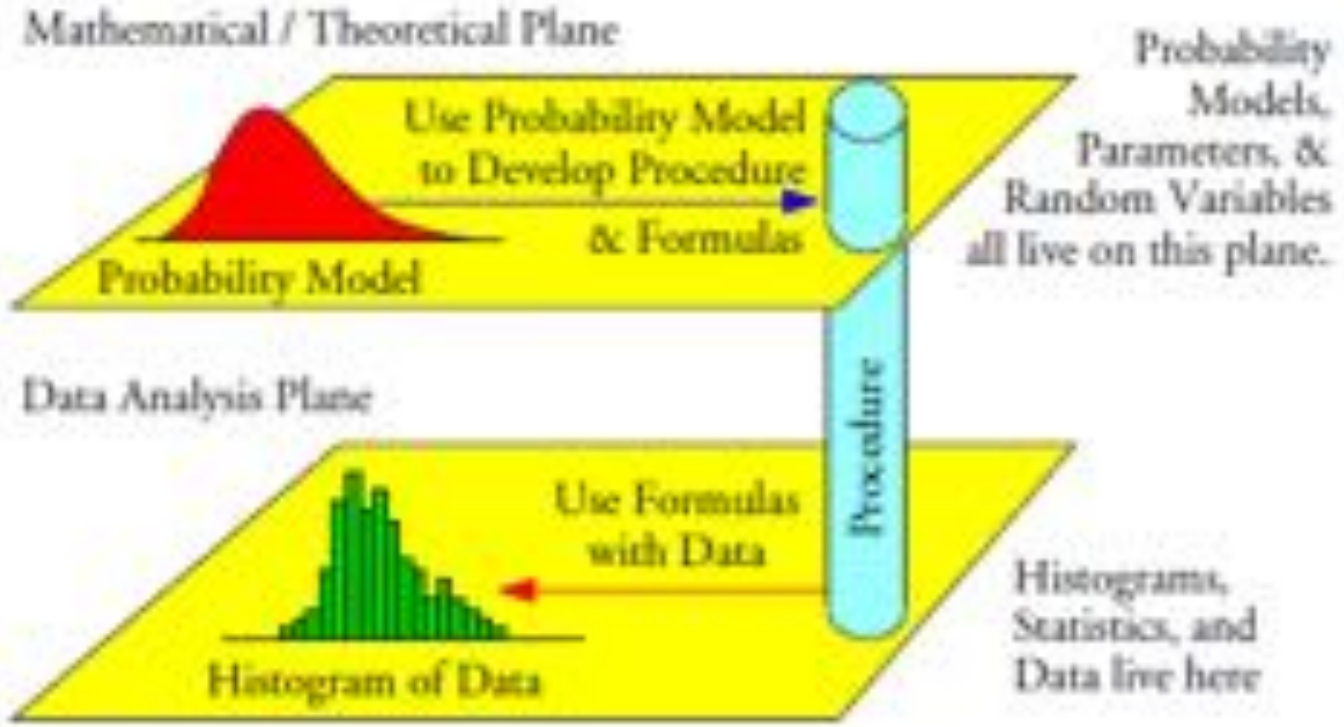
- 动态规划”（dynamic programming）
- “Rao-Blackwell”定理

史春奇

概率 VS 统计

- 理论是概率的， 活还得统计来干！



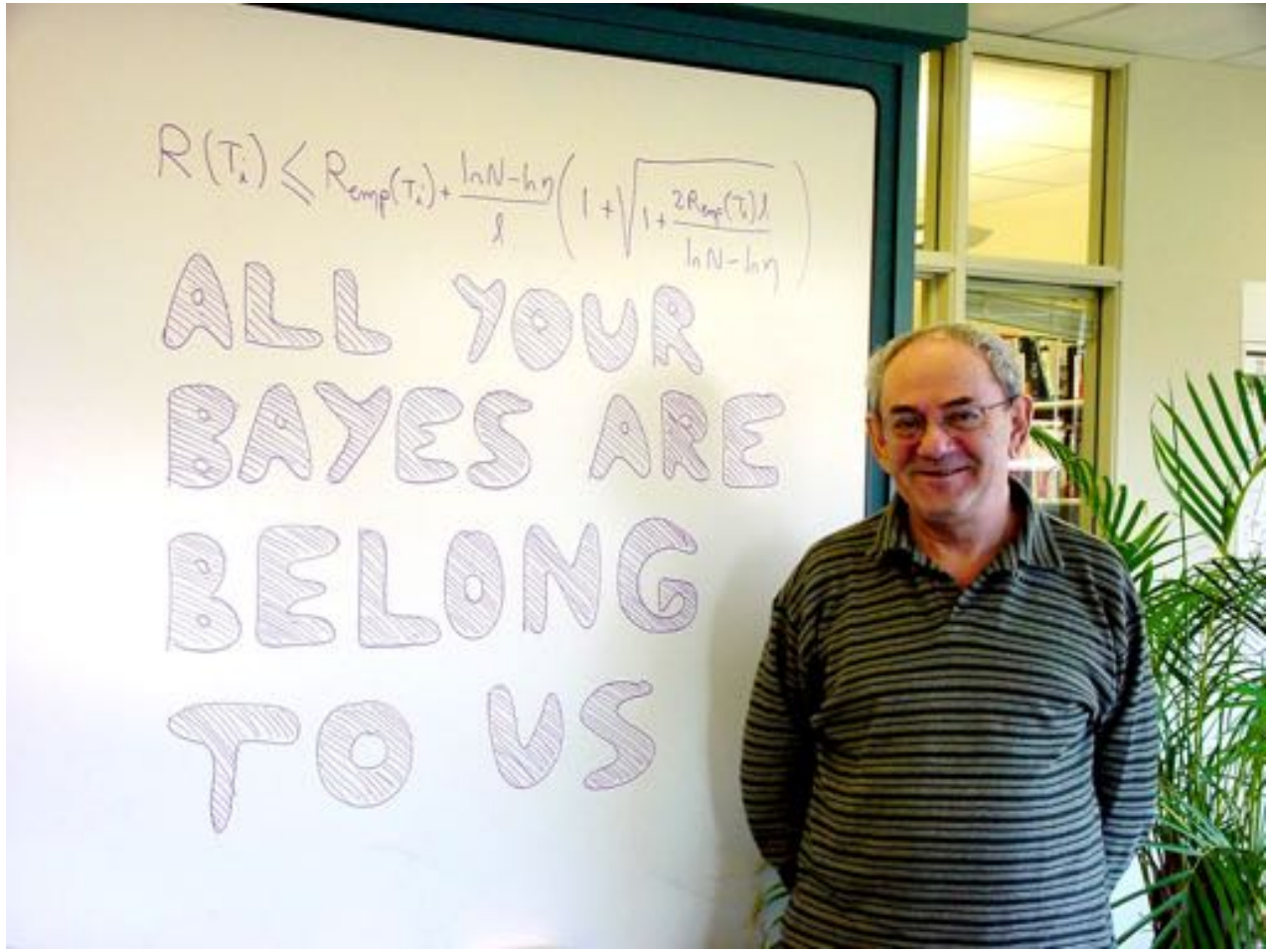


HOW?

- 统计学习大厦如何奠定？

万普尼克 – 苏联黑科技代表

- Vladimir Vapnik
  - Vapnik–Chervonenkis theory
  - Vapnik–Chervonenkis dimension
  - Statistical learning theory
  - Structural risk minimization
  - Support vector machine
- 简历
  - 1958 Uzbek State Univ. 数学硕士
  - 1964 控制科学所 统计系博士
  - 1961 – 1990 科学所计算机系主任
  - 1990 AT&T Bell Labs
  - 目前在Facebook AI Research



史春奇

对比 Jordan 和 Breiman

Michael I. Jordan

Professor of EECS and Professor of Statistics, University of California, Berkeley

machine learning, statistics, computational biology, artificial intelligence, optimization

Verified email at cs.berkeley.edu - Homepage

Citation indices	All	Since 2012
Citations	119299	61567
h-index	138	95
i10-index	430	358

Co-authors View all

David Blei, Martin Wainwright, Zoubin Ghahra...

Title 1–20	Cited by	Year
Latent dirichlet allocation DM Blei, AY Ng, MI Jordan Journal of machine Learning research 3 (Jan), 993-1022	19021	2003
On spectral clustering: Analysis and an algorithm AY Ng, MI Jordan, Y Weiss Advances in neural information processing systems, 849-856	5868	2002
Adaptive mixtures of local experts RA Jacobs, MI Jordan, SJ Nowlan, GE Hinton Neural computation 3 (1), 79-87	3531	1991
Hierarchical mixtures of experts and the EM algorithm MI Jordan, RA Jacobs Neural computation 6 (2), 181-214	2980	1994

Leo Breiman 1928-2005

Professor of Statistics, UC Berkeley

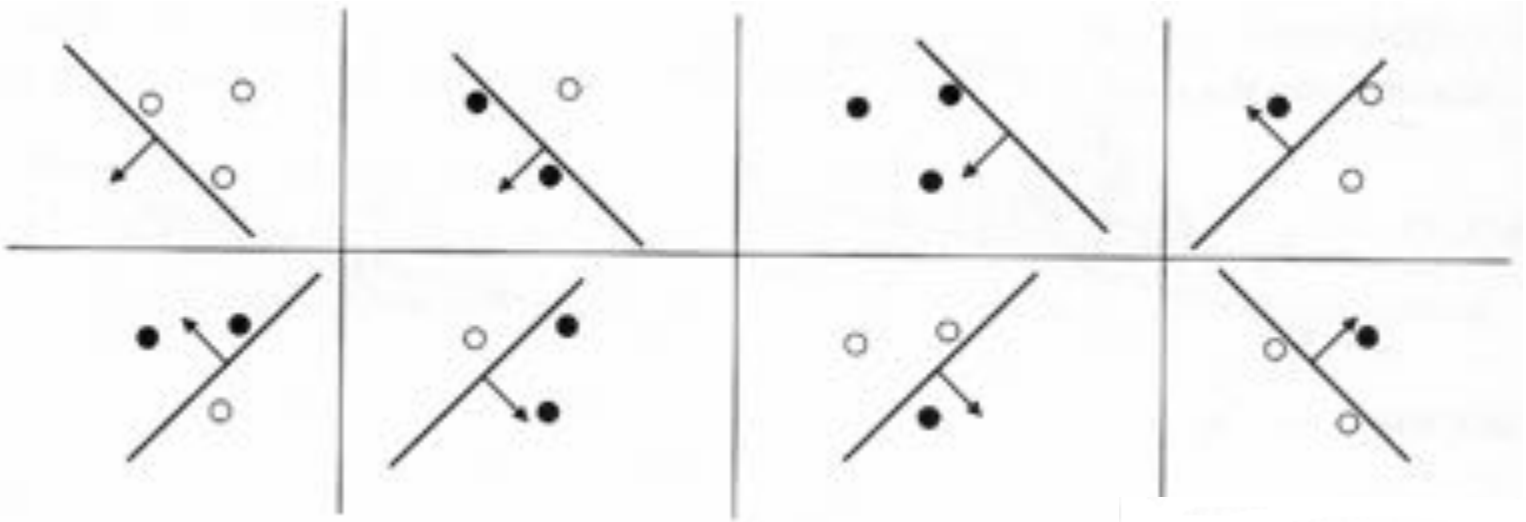
Data Analysis, Statistics, Machine Learning

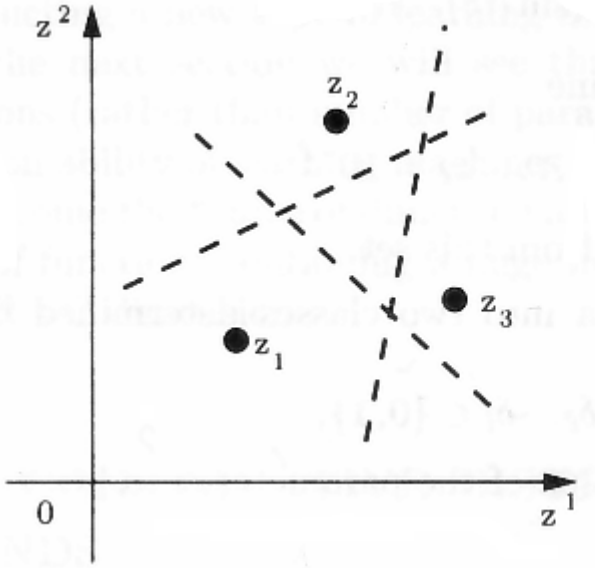
Verified email at stat.berkeley.edu - Homepage

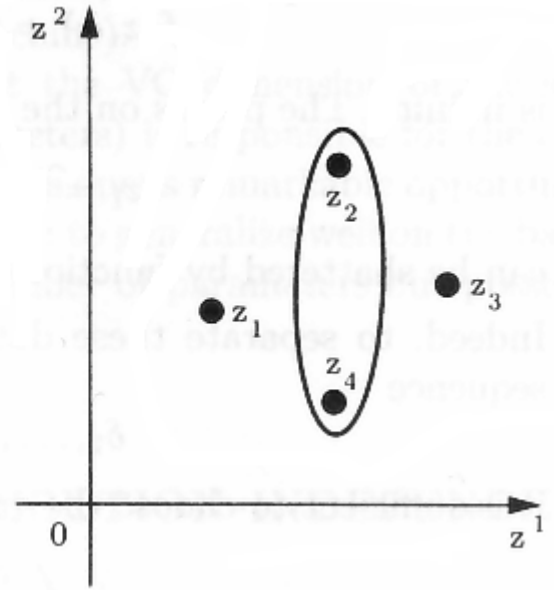
Citation indices	All	Since 2012
Citations	73095	48249
h-index	49	32
i10-index	79	47

Title 1–20	Cited by	Year
Classification and Regression Trees L Breiman, JH Friedman, RA Olshen, CJ Stone CRC Press, New York	32819 *	1999
Classification and regression trees L Breiman Chapman & Hall/CRC	32819	1984
Random forests L Breiman Machine learning 45 (1), 5-32	28809	2001
Bagging predictors L Breiman Machine learning 24 (2), 123-140	16710	1996

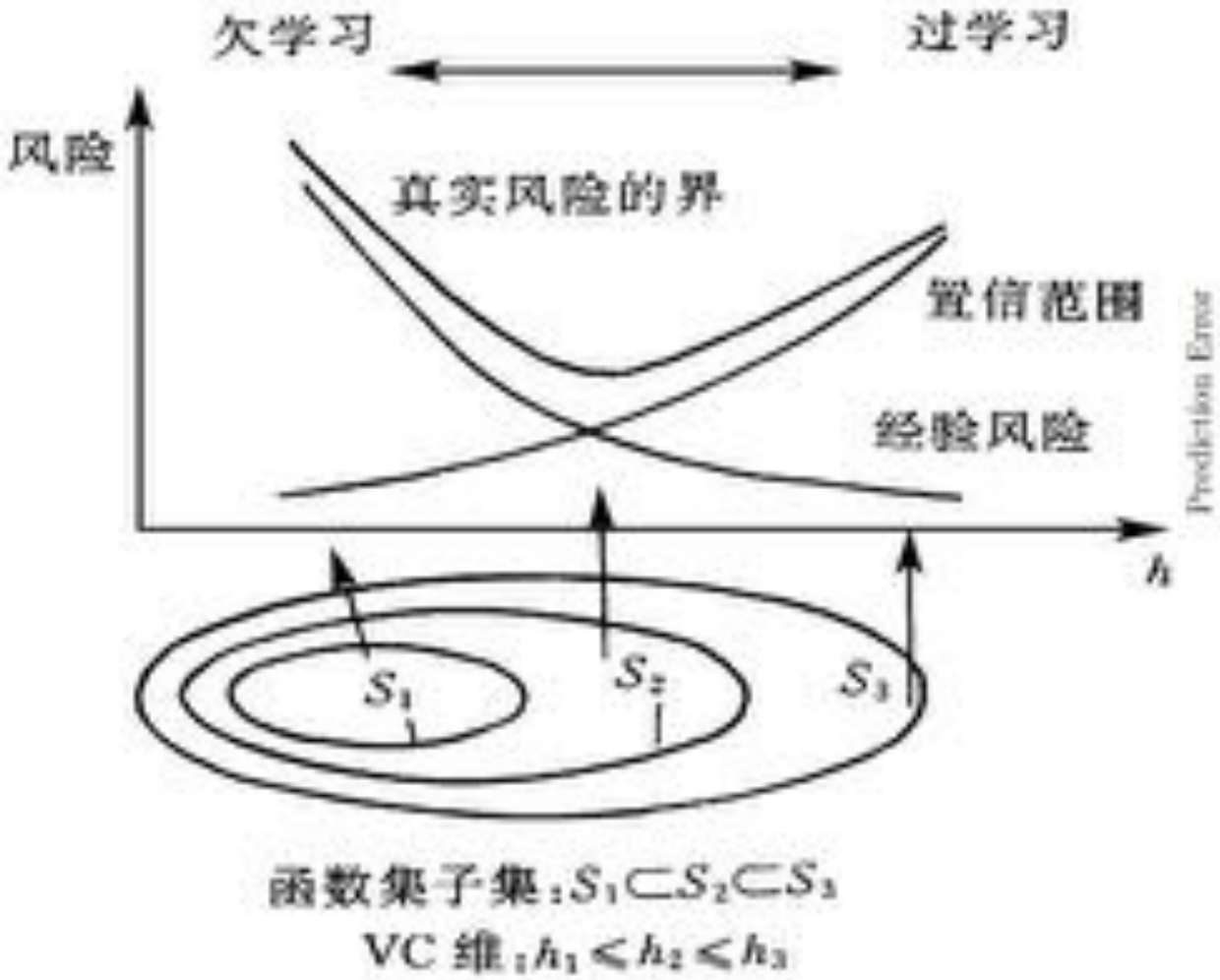
VC维

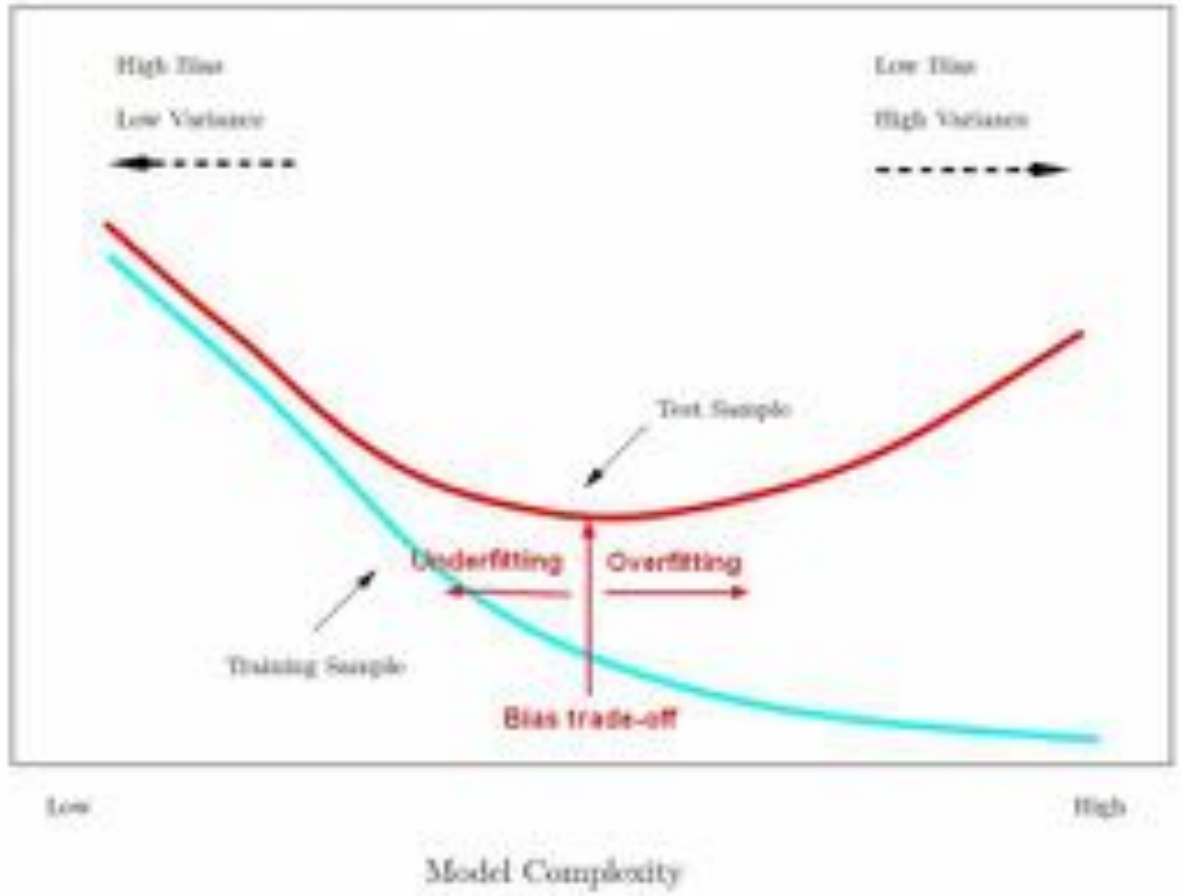




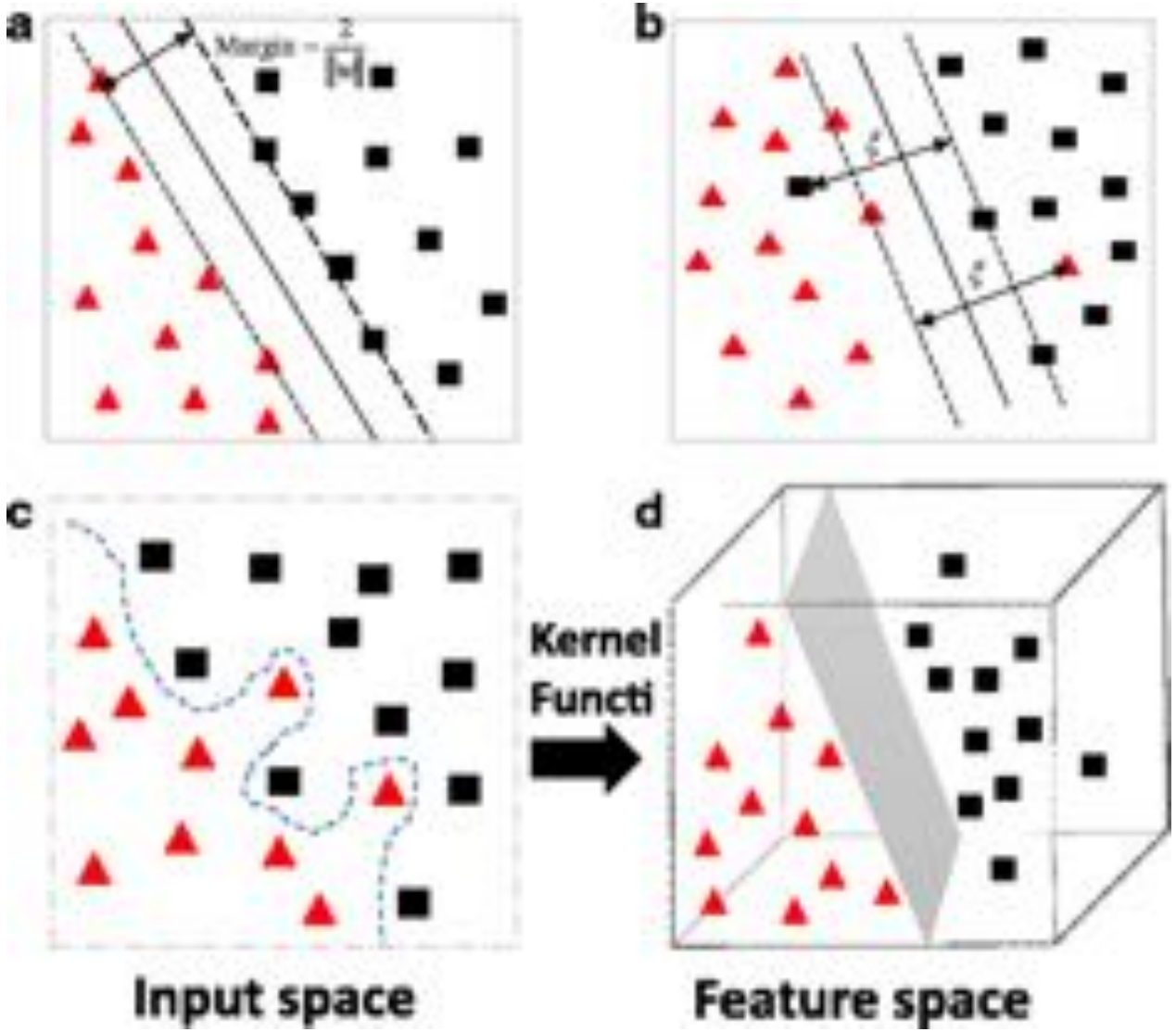


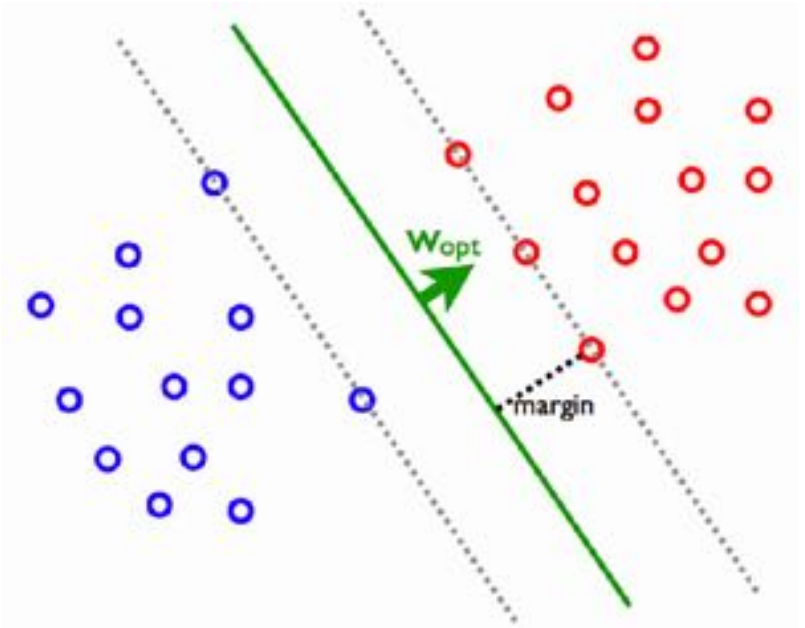
结构风险最小 SRM


欠学习
过学习
风险
真实风险的界
置信范围
经验风险
h
S_1
S_2
S_3
函数集子集：$S_1 \subset S_2 \subset S_3$
VC 维：$h_1 \leqslant h_2 \leqslant h_3$


High Bias
Low Variance
Low Bias
High Variance
Prediction Error
Test Sample
Underfitting
Overfitting
Training Sample
Bias trade-off
Low
High
Model Complexity

支持向量机 Support Vector Machine, SVM

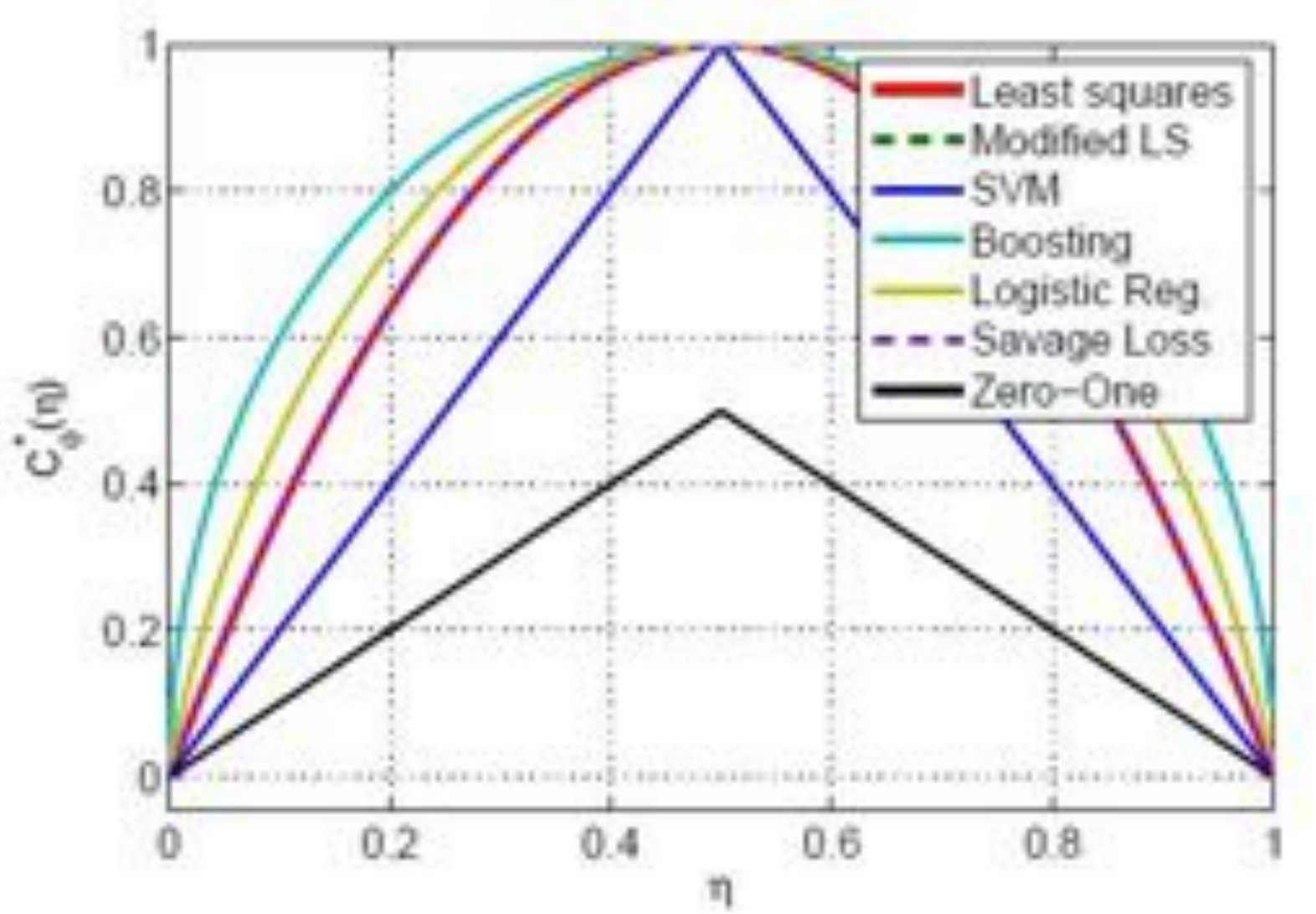




风险和边缘理论 统一 分类学习

史春奇

- Risk & Margin



WHY RUSSIAN?

- 为什么是俄国人？

俄国数学崛起

欧拉 在 圣彼得堡

- 俄国皇家科学院
  - 欧拉娶了科学院附属中学的美术教师

叶卡捷琳娜一世 (凯萨琳一世)

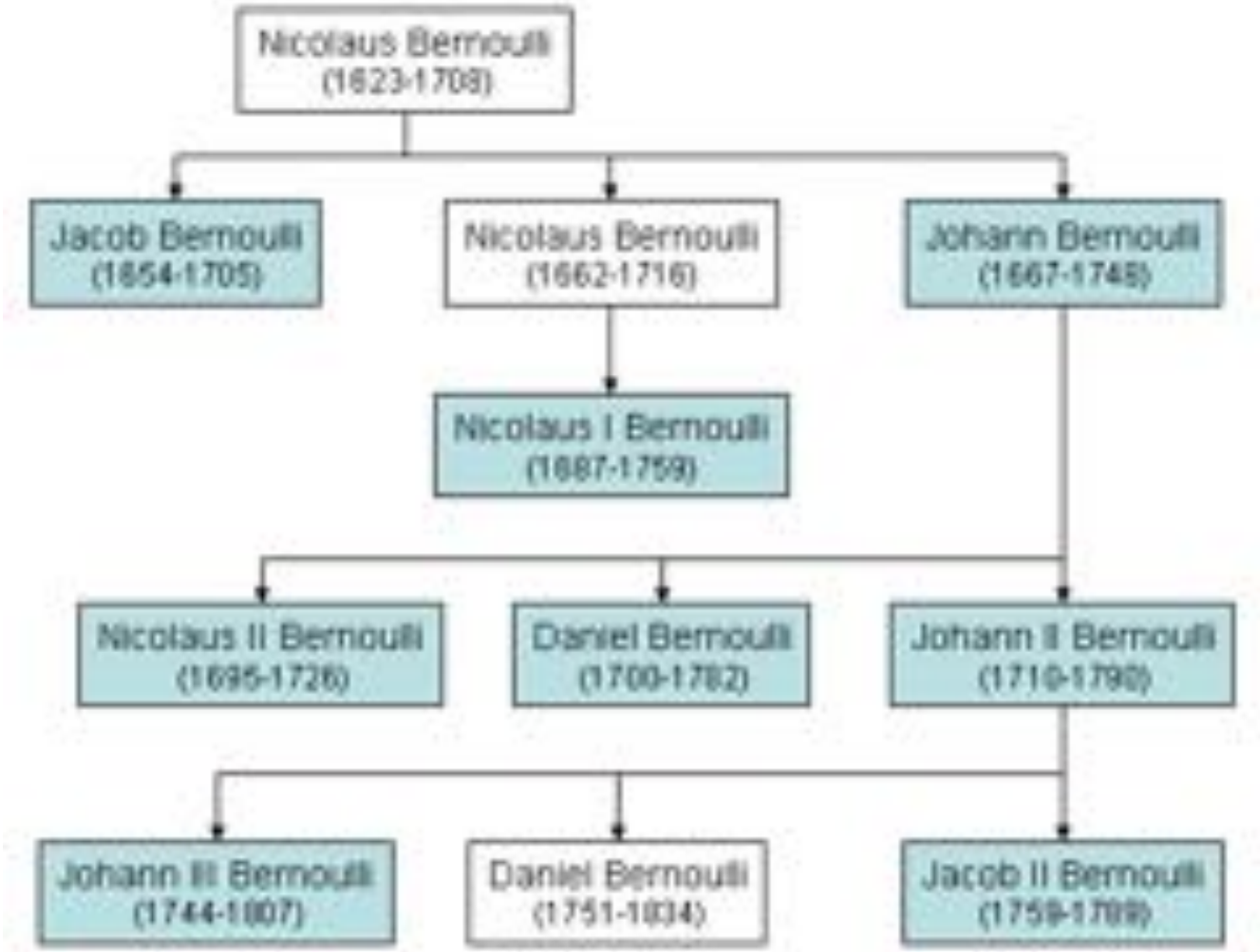
伯努利

雅各布·伯努利

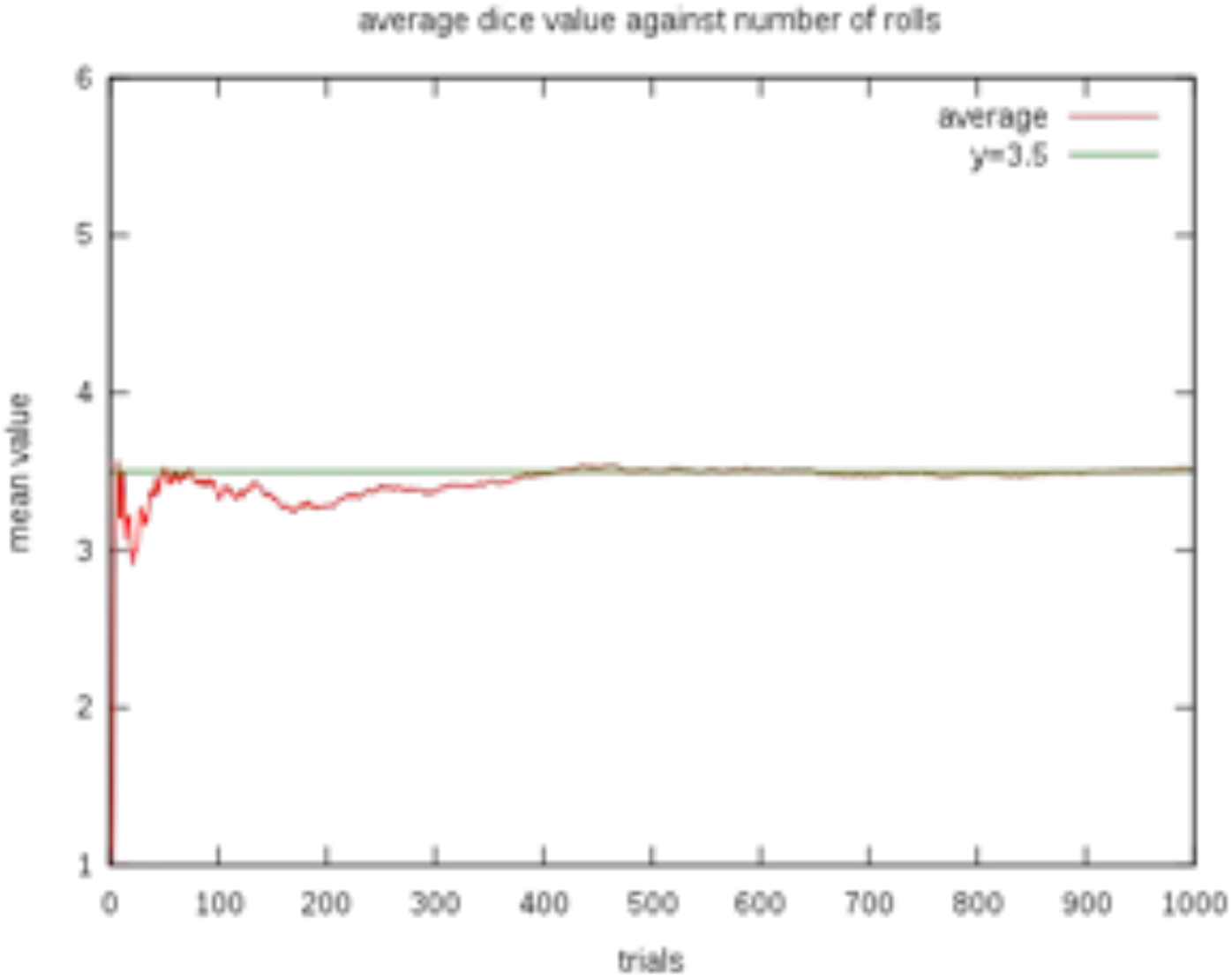
约翰·伯努利

欧拉的博士导师

伯努利家族


Nicolaus Bernoulli
(1623-1708)
Jacob Bernoulli
(1654-1705)
Nicolaus Bernoulli
(1662-1716)
Johann Bernoulli
(1667-1748)
Nicolaus I Bernoulli
(1687-1759)
Nicolaus II Bernoulli
(1695-1726)
Daniel Bernoulli
(1700-1782)
Johann II Bernoulli
(1710-1790)
Johann III Bernoulli
(1744-1807)
Daniel Bernoulli
(1751-1834)
Jacob II Bernoulli
(1759-1789)

伯努利大数定律



$$\lim_{n\to+\infty} P\left\{\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} E(X_i)\right| < \varepsilon\right\} = 1 \cdots\cdots (1)$$

$$\lim_{n\to+\infty} P\left\{\left|\frac{\mu_n}{n} - \frac{1}{n}p\right| < \varepsilon\right\} = 1 \cdots\cdots (2)$$

$$\frac{1}{n^2}\left(\sum_{i=1}^{n} X_i\right) \to 0 \cdots\cdots (3)$$

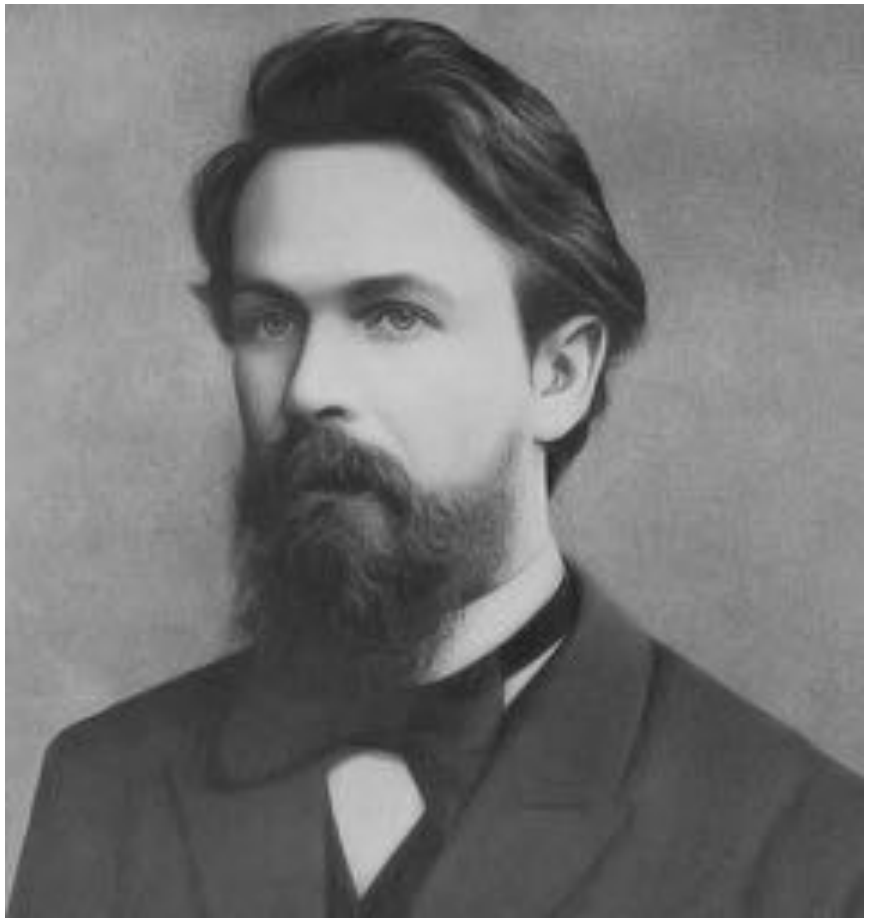
切比雪夫 – 彼得堡学派奠基人

- 1821年 – 1894年
- 法语通信
- 研究欧拉手稿
- 概率不等式研究
- 培养学生
  - 马尔科夫 Markov
  - 李雅普诺夫 Lyapunov
- 圣彼得堡国立大学

$$\Pr(|X - \mathbf{E}(X)| \geq a\,\sigma) \leq \frac{1}{a^2}$$

发扬光大

马尔科夫 Markov



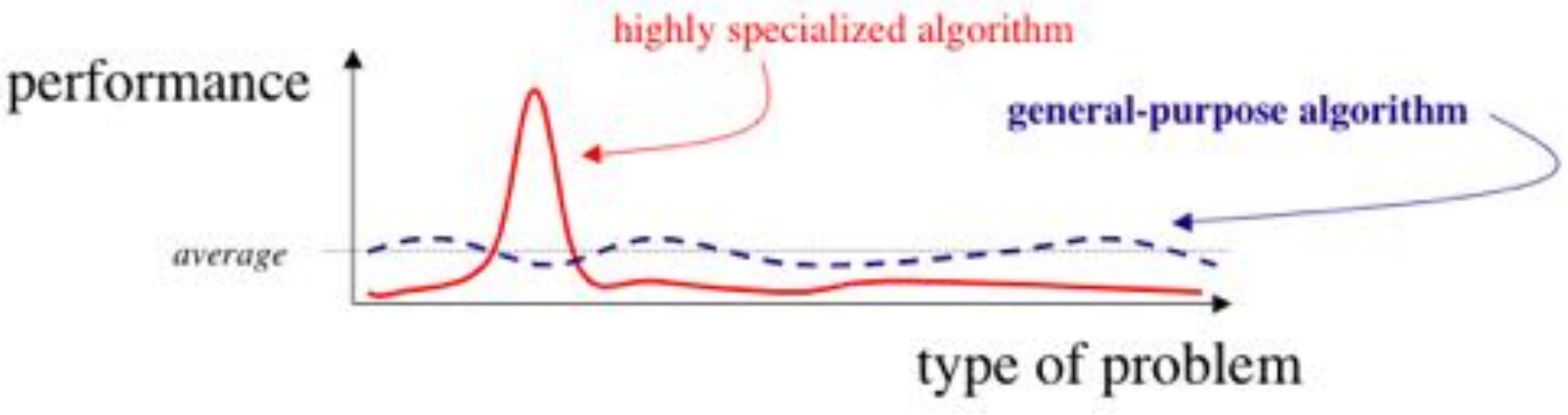
李雅普诺夫 Lyapunov

ENGINE?

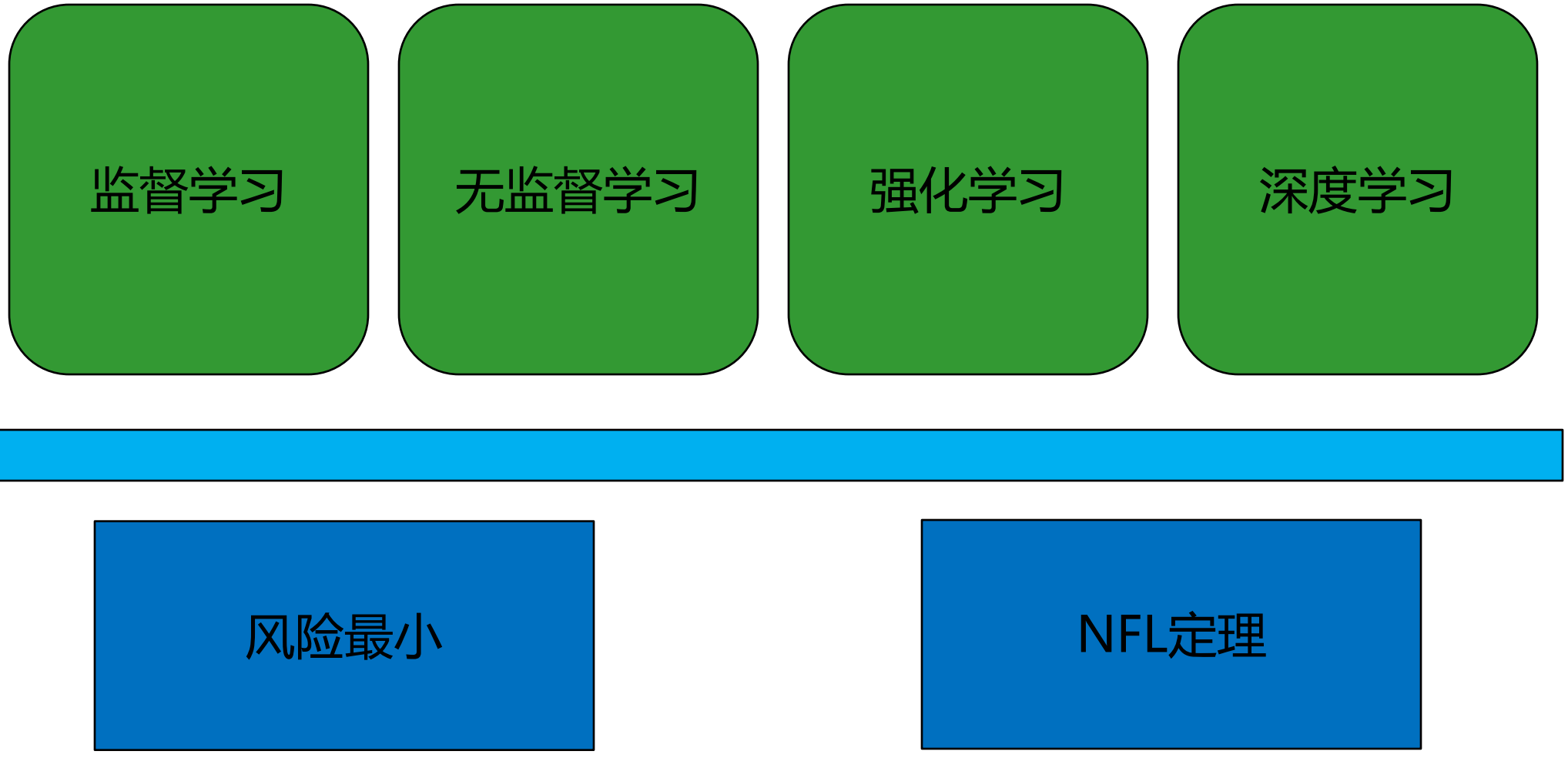
- 除了框架还有优化！

NFL 天下没有免费的午餐定理

- No free lunch theorem， NFL
- David Wolpert 提出的优化理论
- 集成学习Ensemble learning#Stacking 优化



机器学习的大厦



类比金融数学理论

史春奇

衍生物定价
Derivatives Pricing

风险分析
Risk Analysis and Reduction

组合优化
Portfolio Optimization

统计套利
Statistical Arbitrage

市场效率原理
Principle of market efficiency

无套利原理
Principle of no arbitrage

MORE?

- 统计学习进一步发展？

Leo Breiman

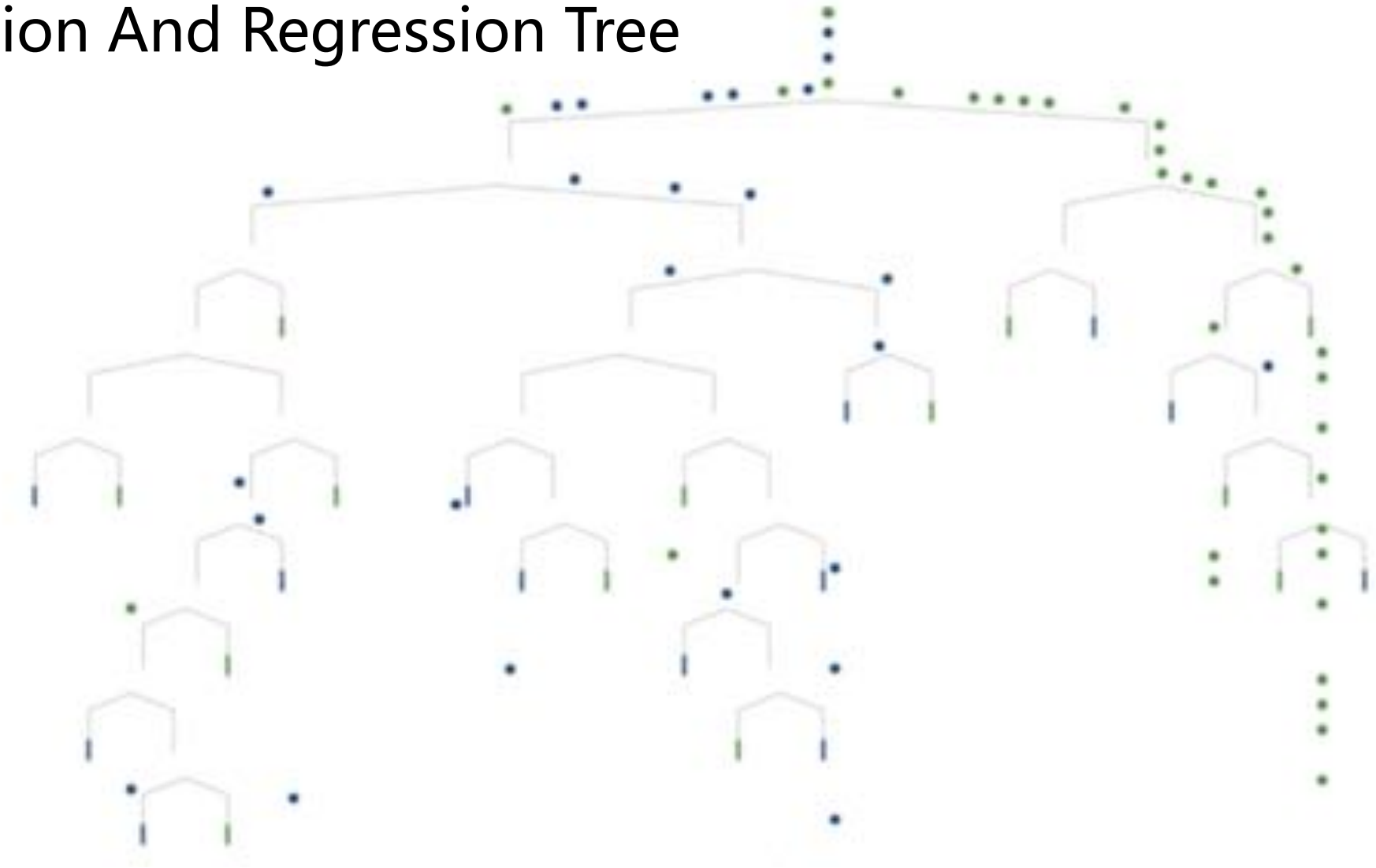
- 1928年 - 2005年
- 加州理工Caltech物理系本科
- 哥伦比亚大学数学系硕士
- 伯克利统计系博士
- 导师Michel Loève是创系10元老之一

- 填补计算机和统计之间鸿沟的先驱！

- Shannon-McMillan-Breiman(SMB)定理
- CART, Bagging, Random forest

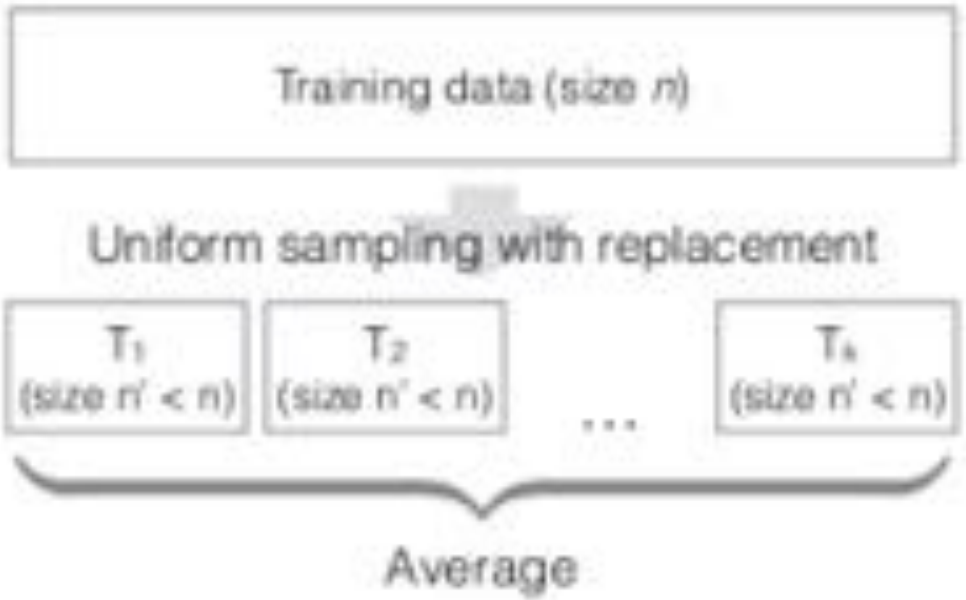
Decision Tree: CART 1984

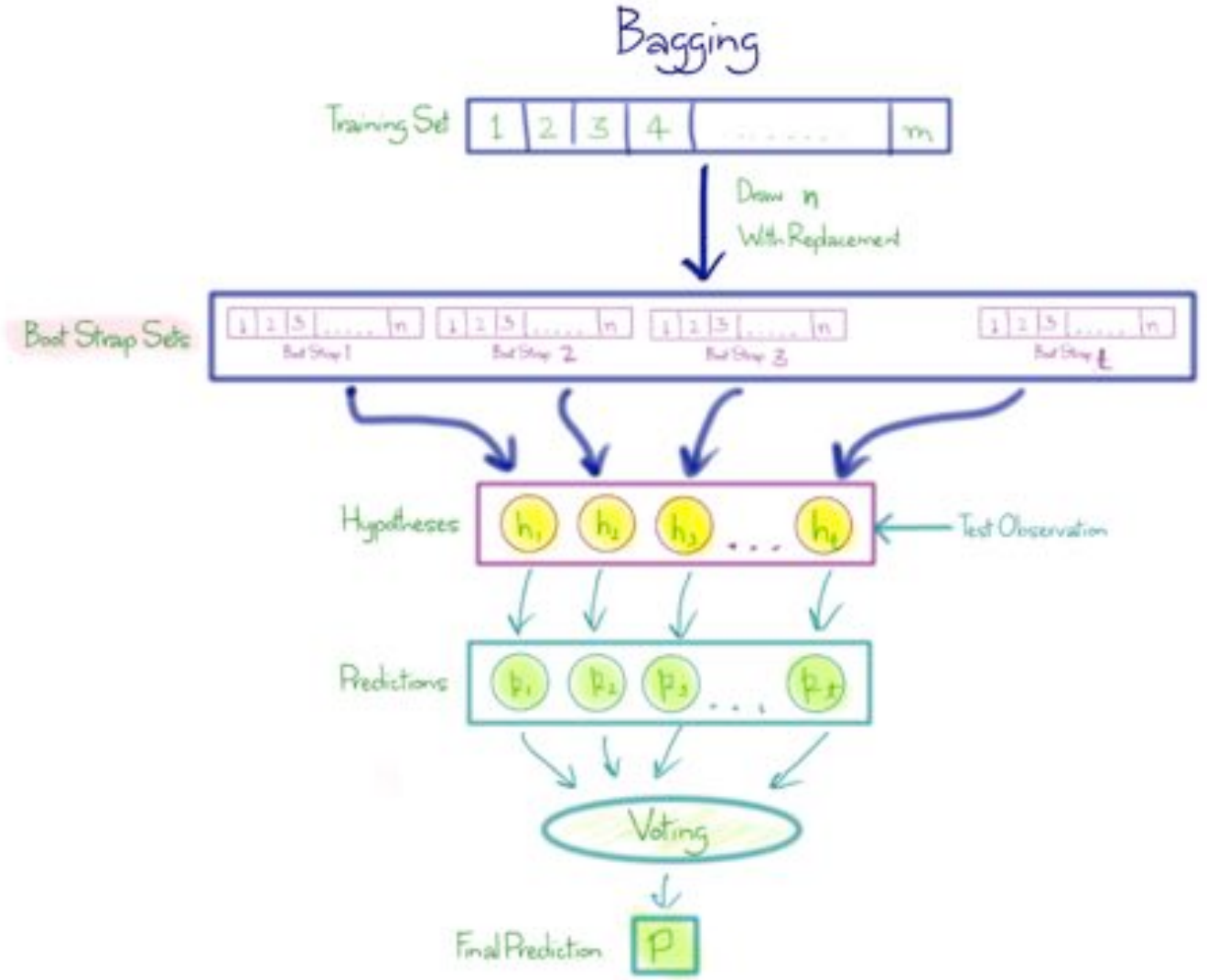
- Classification And Regression Tree



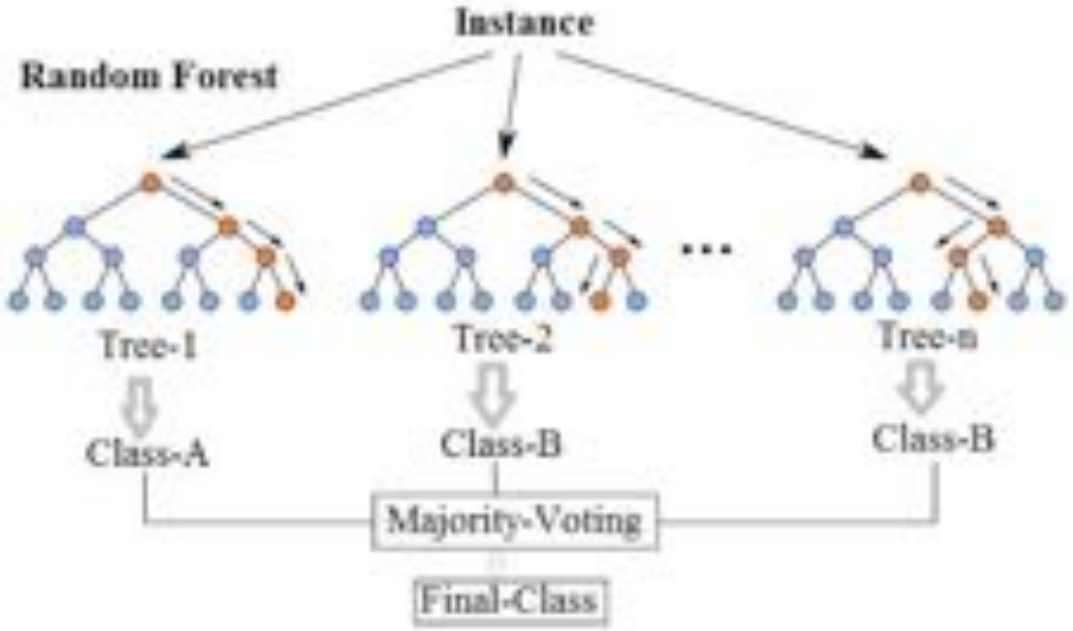
Bagging 1994

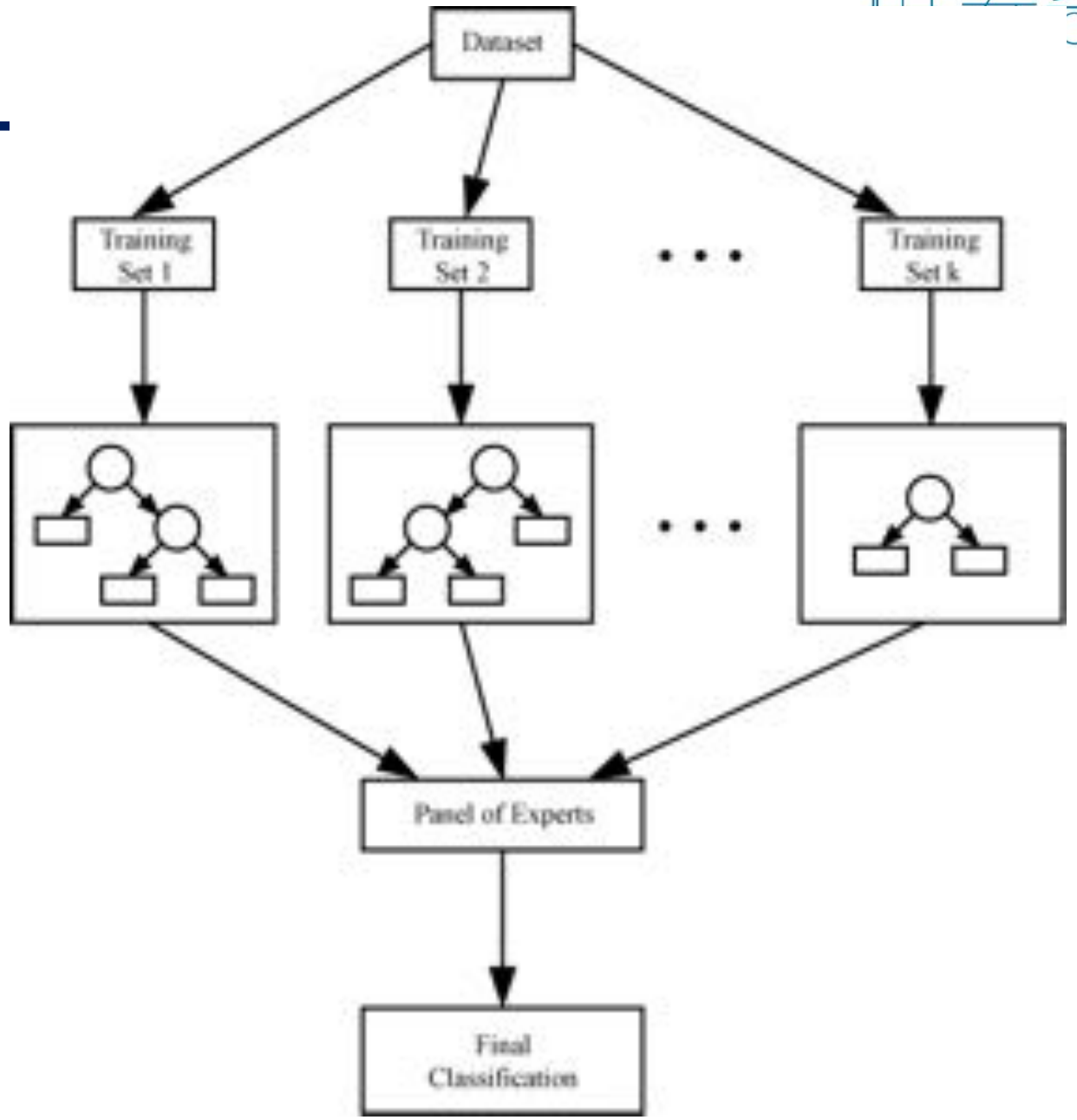
- Bootstrap Aggregating





随机森林 Random Forest 2001



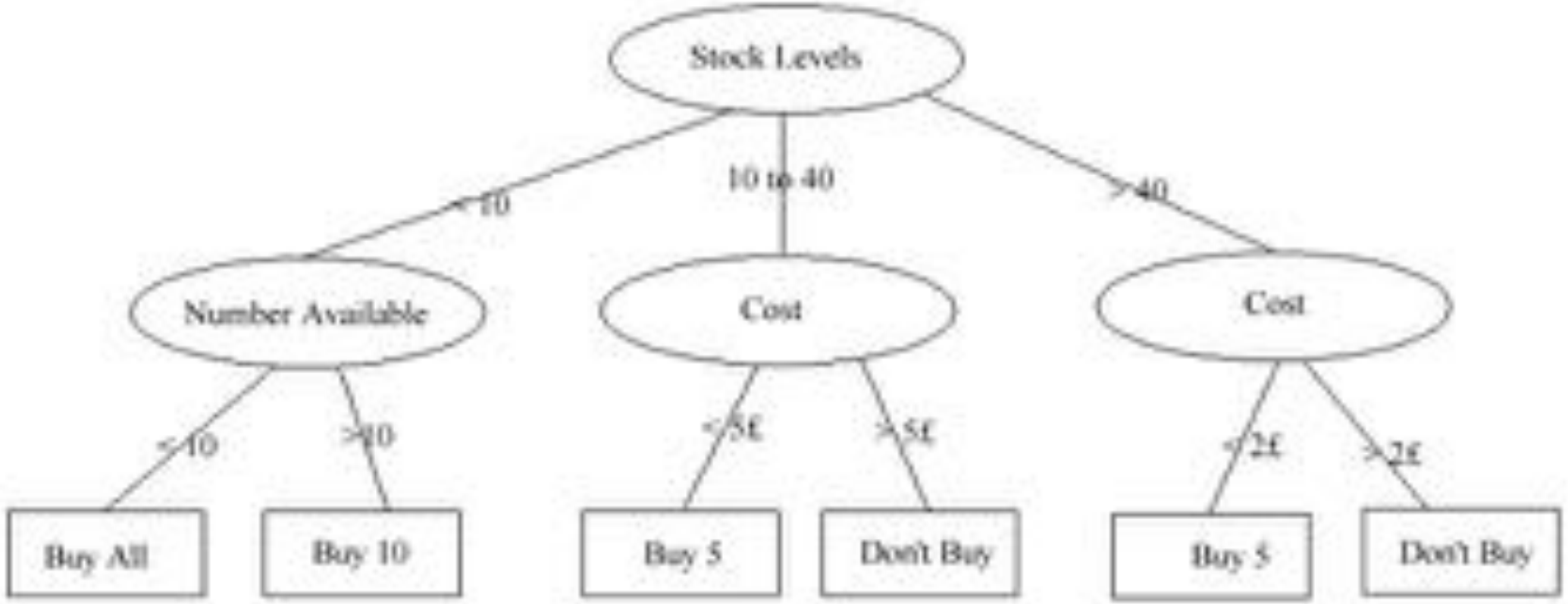


Ross Quinlan

- University of Sydney 物理计算机双本科
- University of Washington计算机博士
- University of Sydney教授

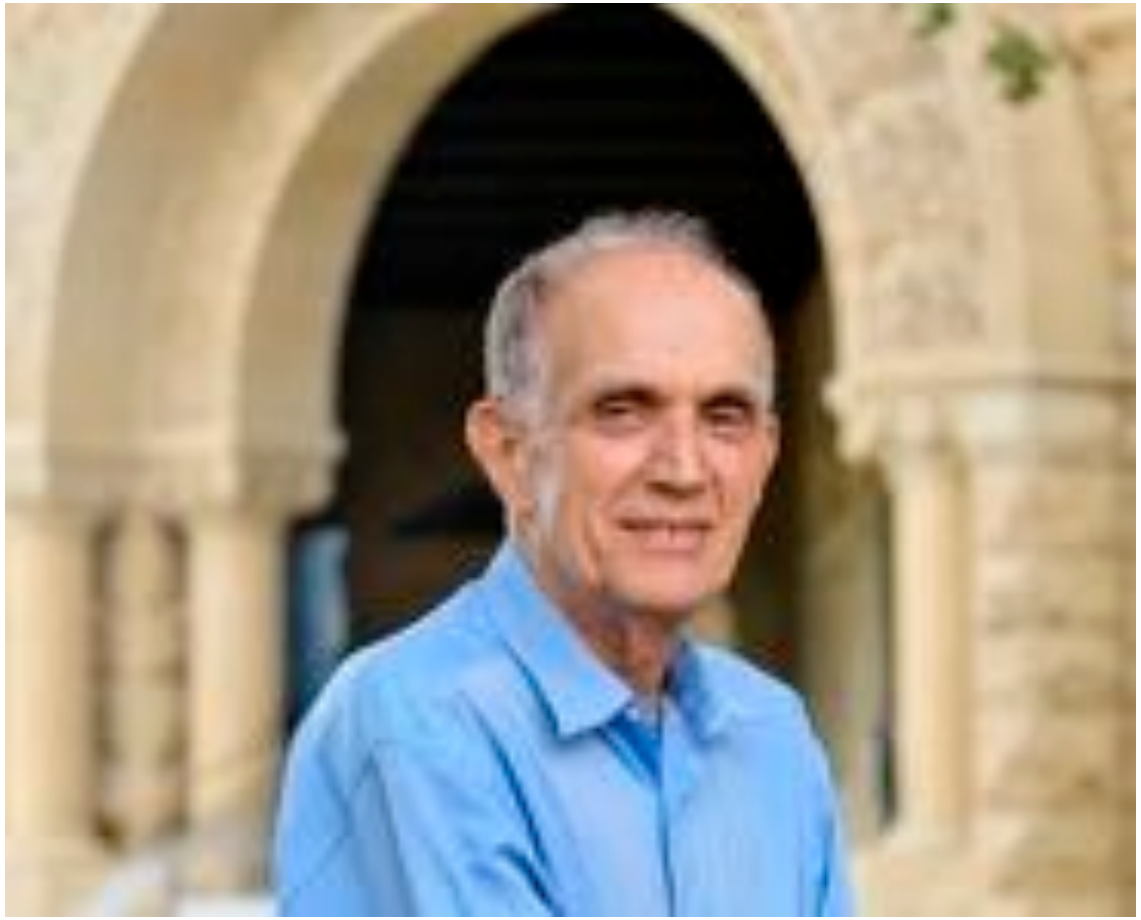
- ID3, C4.5, C5.0

C4.5 决策树


Stock Levels
<10
10 to 40
>40
Number Available
Cost
Cost
<10
>10
<5£
>5£
<2£
>2£
Buy All
Buy 10
Buy 5
Don't Buy
Buy 5
Don't Buy

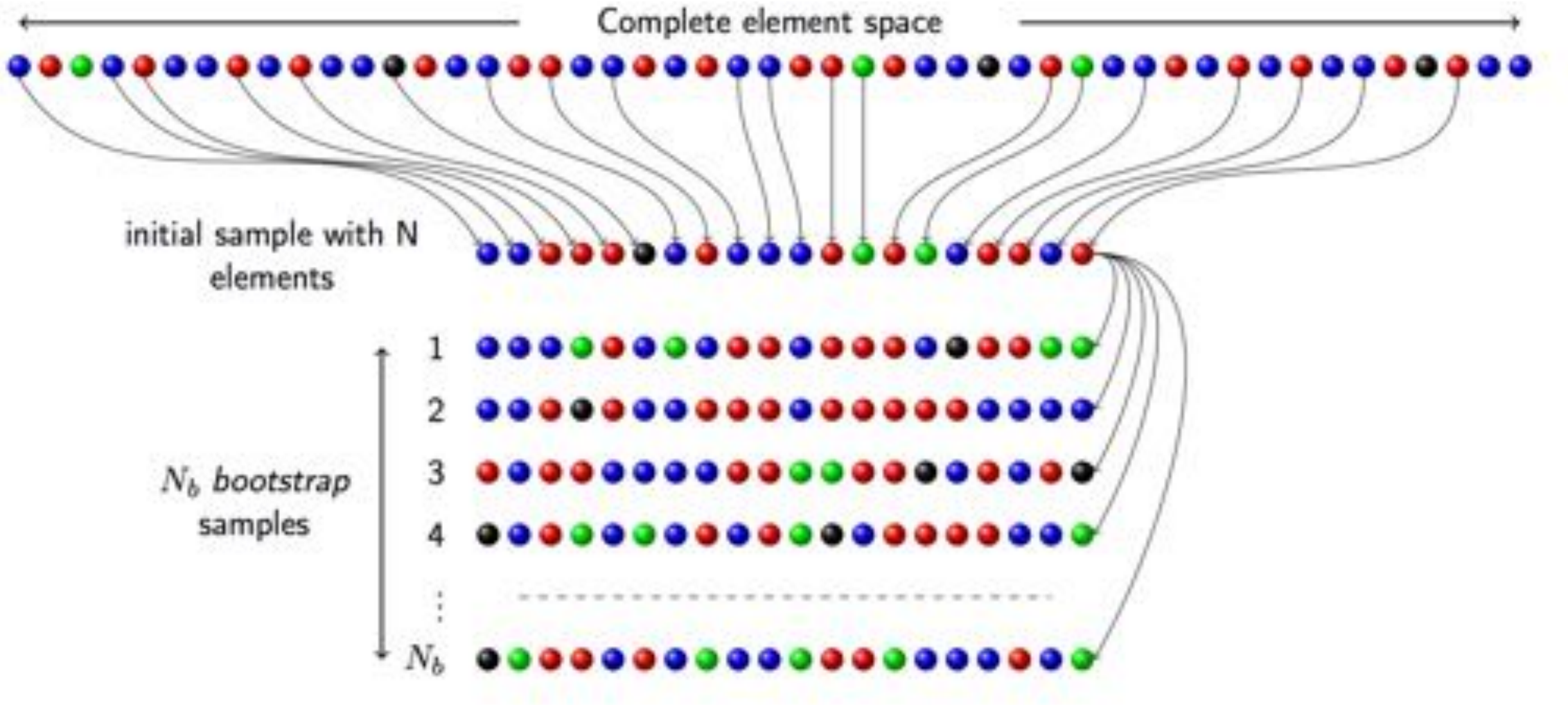
Bradley Efron

- 加州理工Caltech物理系本科
- 斯坦福博士

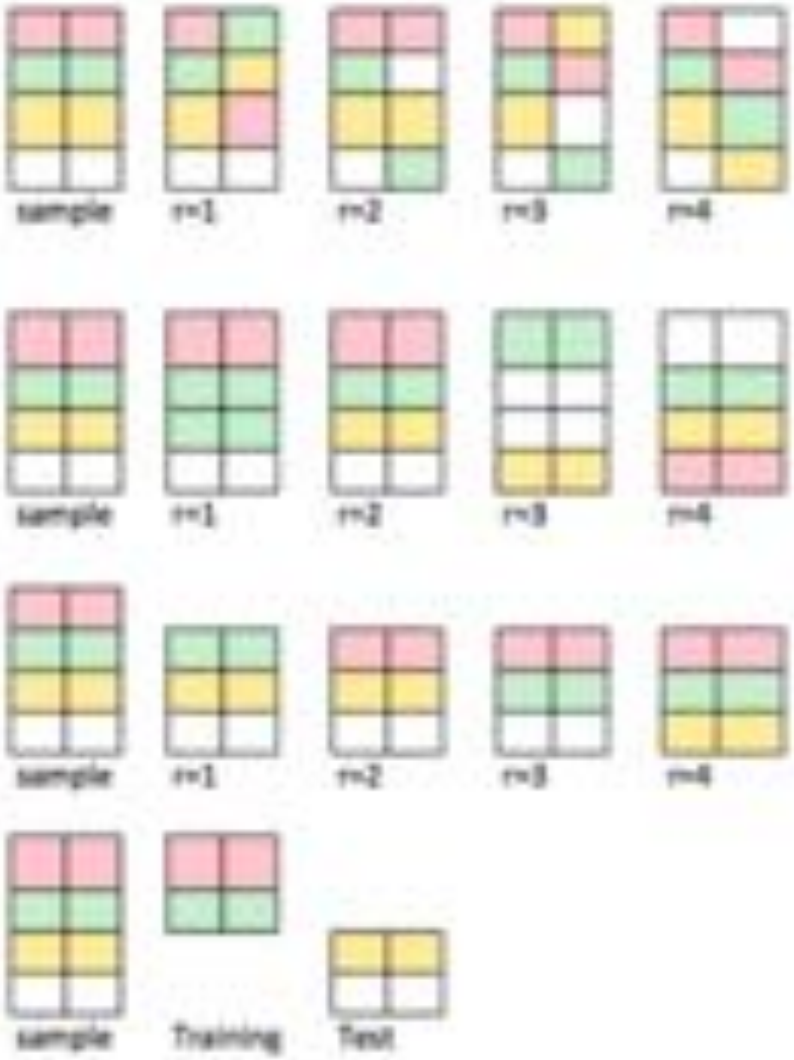


- Bootstrap 采样
- LAR 回归

Bootstrap 采样



Bootstrap VS Jacknife



Permutation
Randomization test

费希尔 Fisher

Bootstrap

Bradley Efron

Jacknife

John W. Tukey

cross validation

John W. Tukey

史春奇

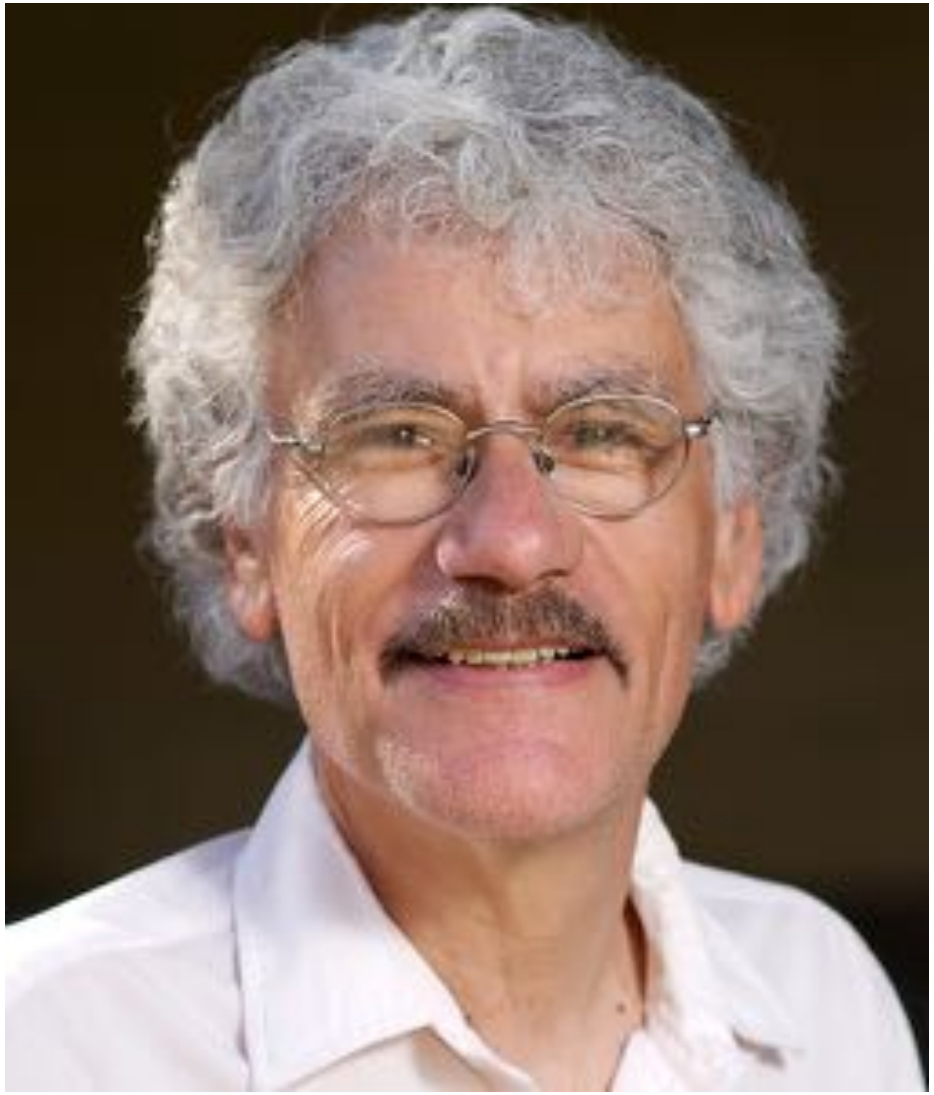
- 1915年-2000年
- Princeton University 数学博士

- 学生
  - 钟开莱Kai Lai Chung

- FFT algorithm
- Tukey's range test
- Box plot

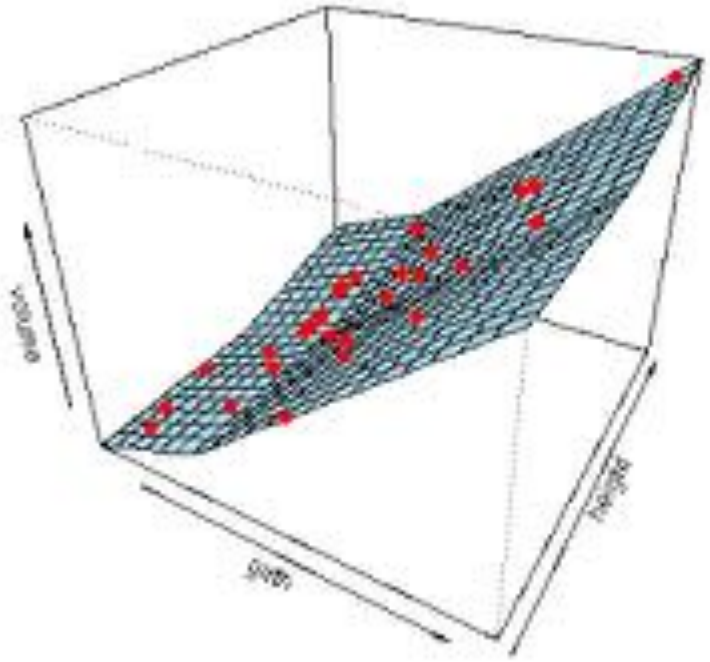
Jerome H. Friedman

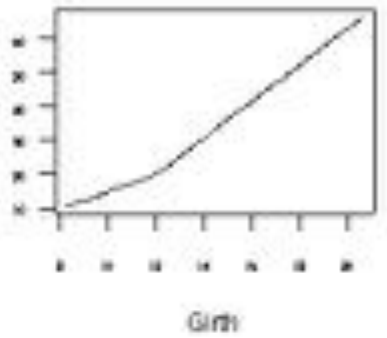
- 伯克利物理系博士
- 斯坦福统计系教授
- Breiman合作发明CART树
- Multivariate Adaptive Regression Splines, MARS
- Gradient boosting算法

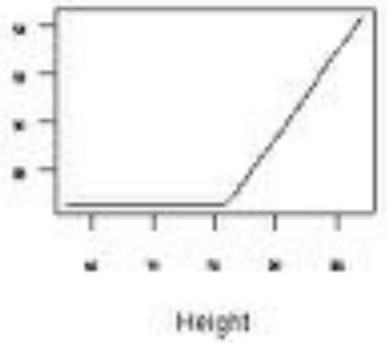


MARS

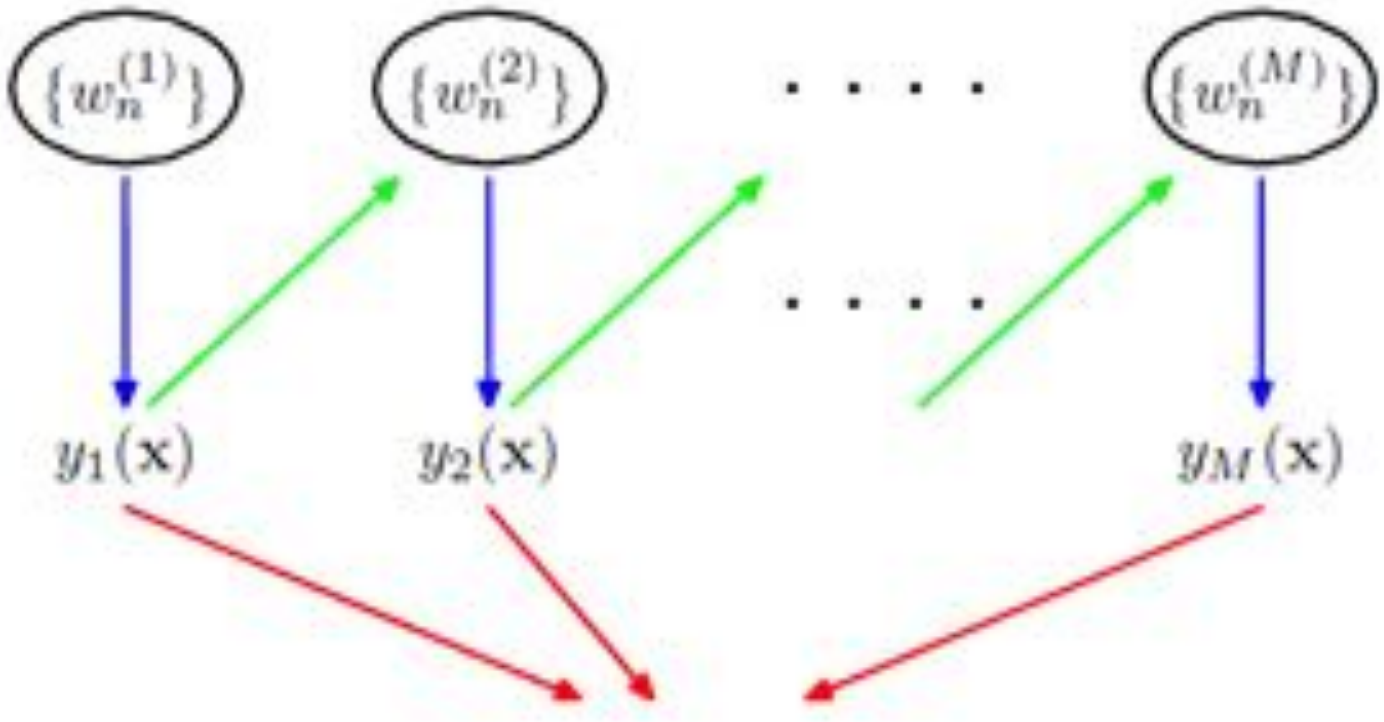
- Multivariate Adaptive Regression Splines







Gradient boosting



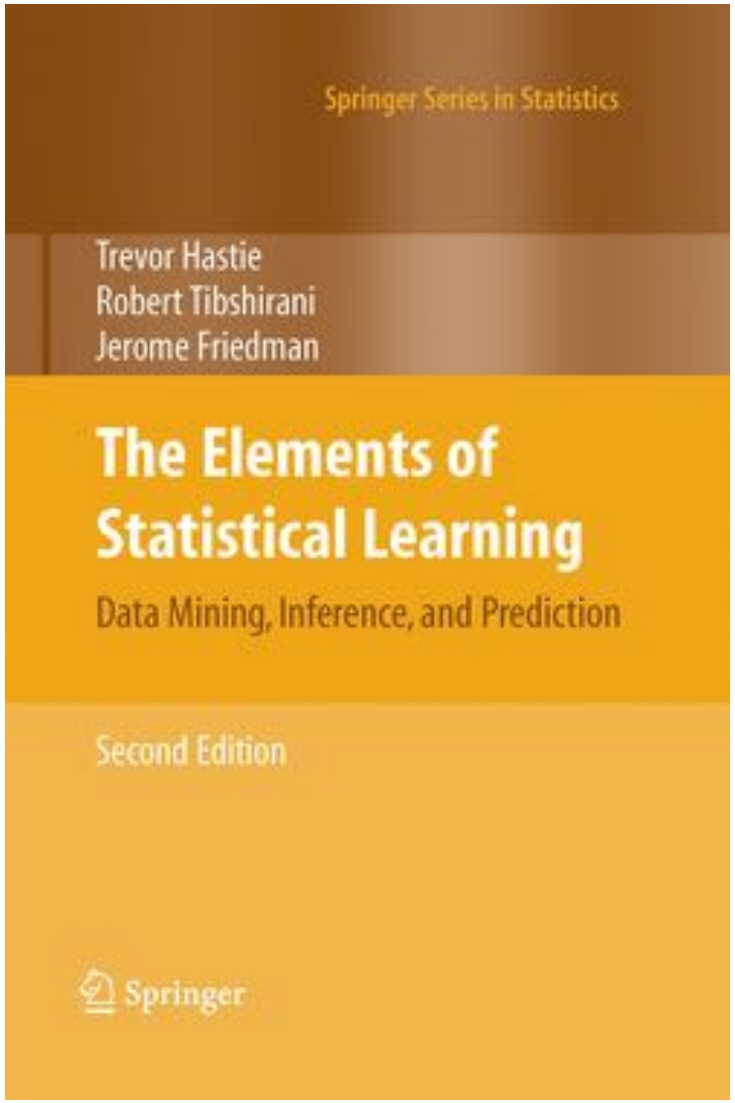
$$Y_M(\mathbf{x}) = \text{sign}\left(\sum_m^M \alpha_m y_m(\mathbf{x})\right)$$

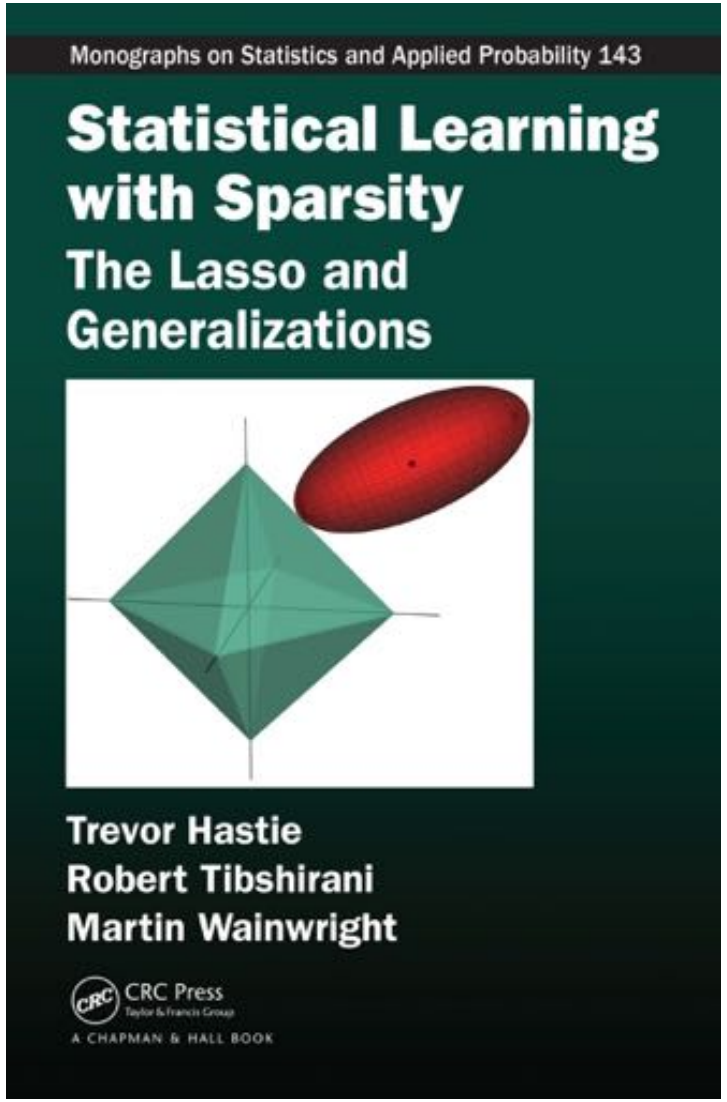
Trevor Hastie

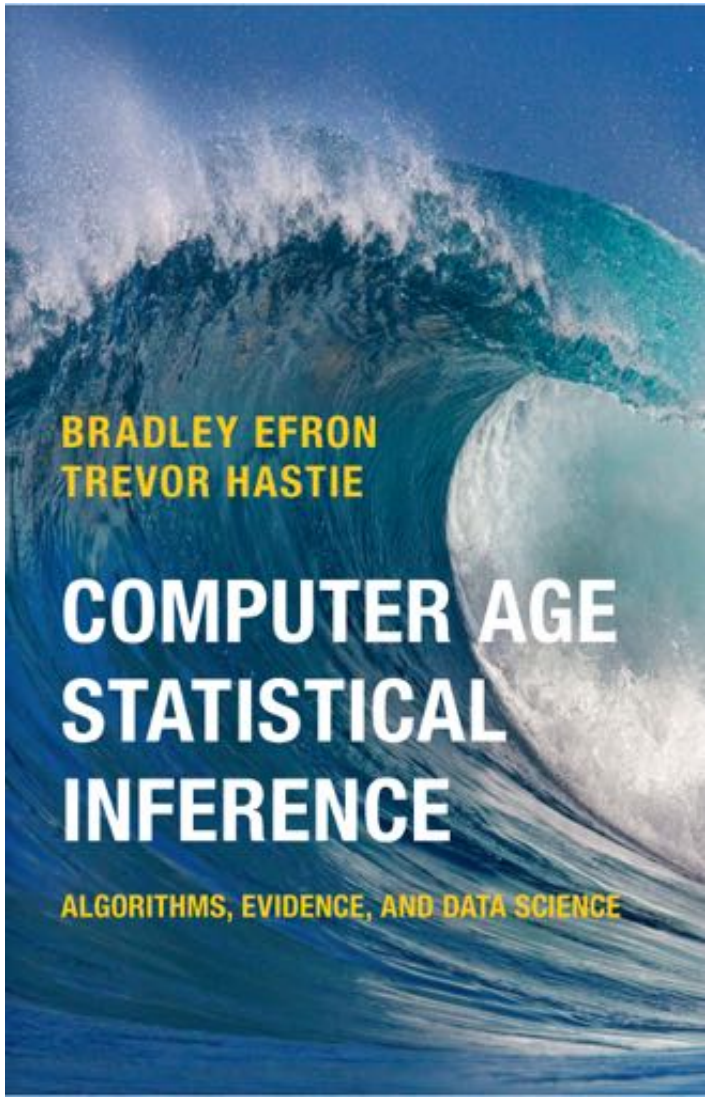
- 斯坦福大学博士
- 斯坦福大学教授
- 标志着斯坦福大学成为另外一个统计牛校

史春奇

Trevor Hastie 三本书







Robert Tibshirani

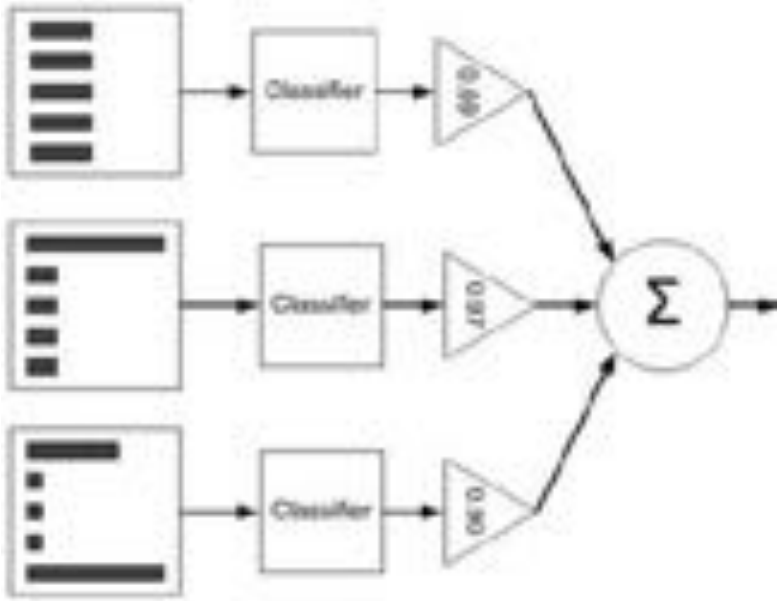
- 斯坦福大学博士
  - Efron的博士弟子
  - Hastie的同学
- 斯坦福大学教授
- LASSO method

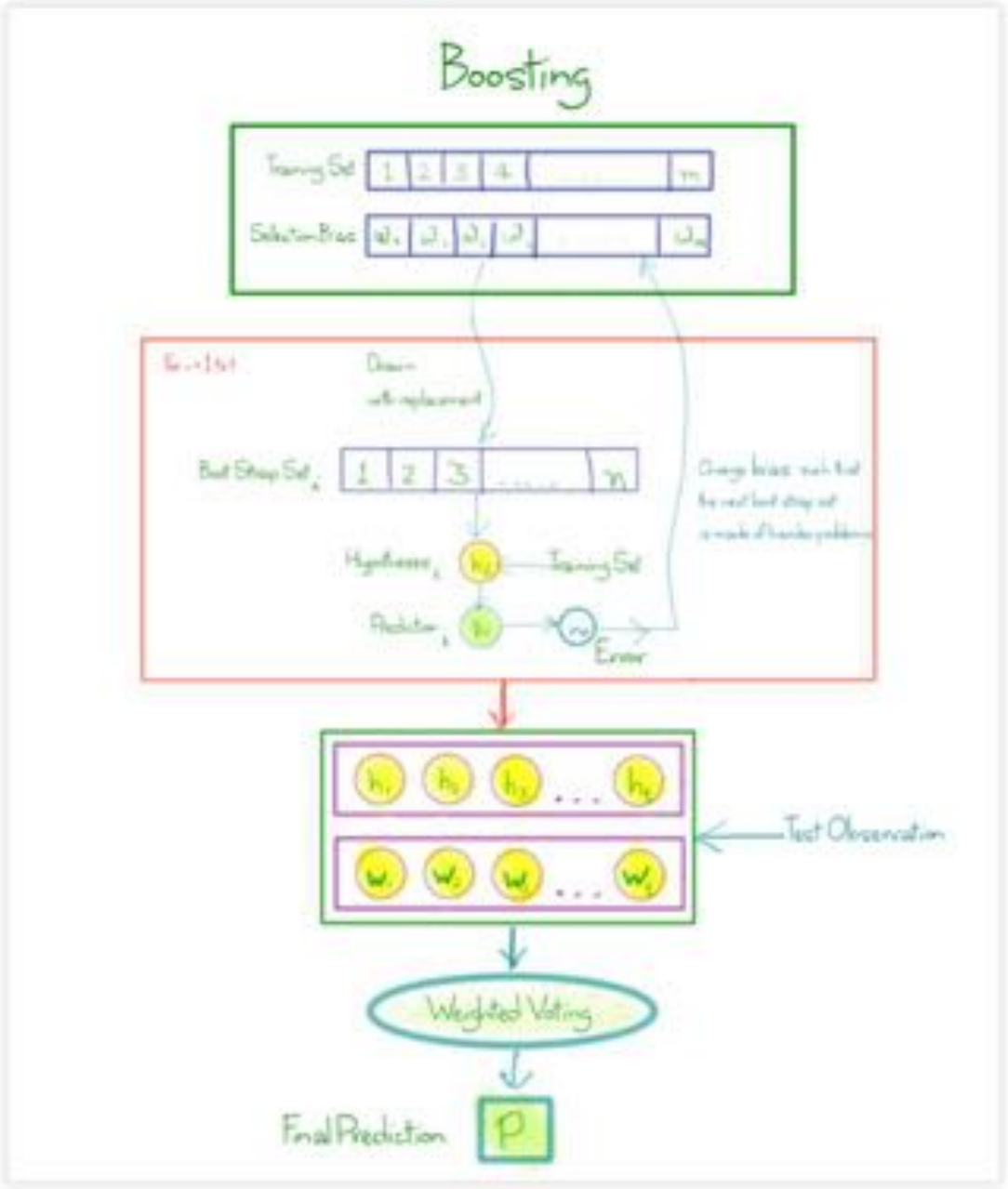
Robert Schapire

- MIT博士
- 微软研究员

- 和Yoav Freund一起发明Adaboost

Boosting

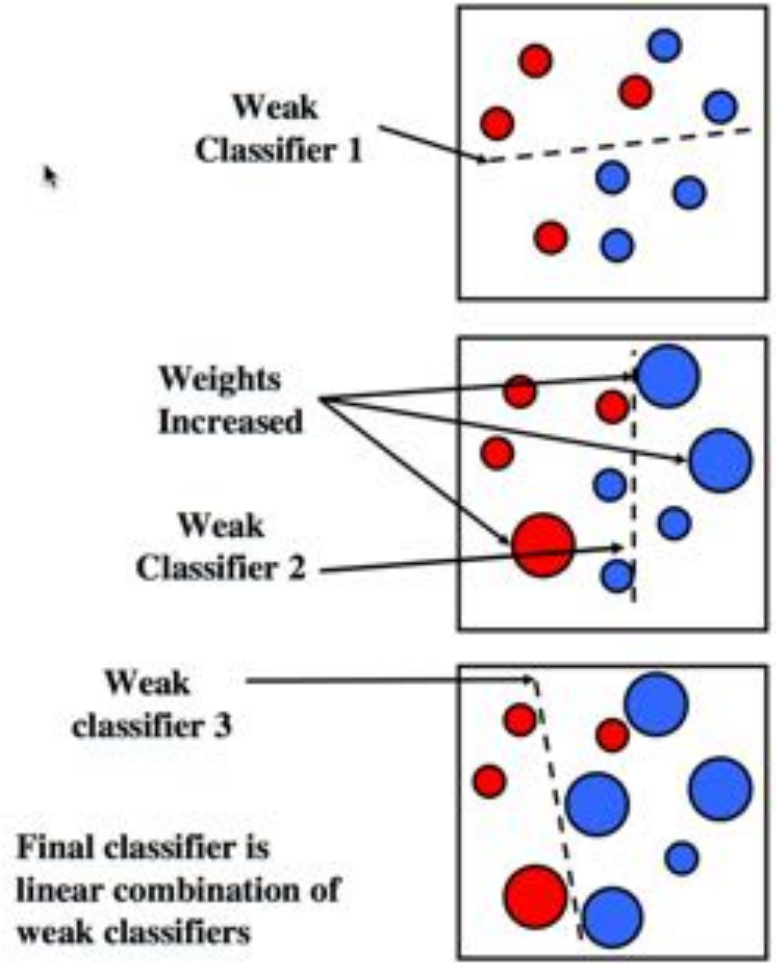

Classifier
Classifier
Classifier
Σ

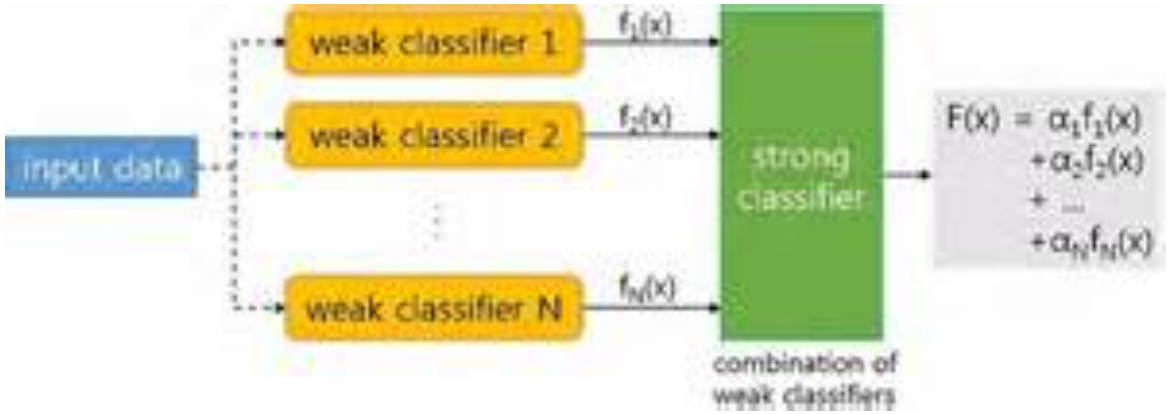

Boosting
Training Set
Selection Bias
Bootstrap Set
Draw with replacement
Hypothesis
Training Set
Prediction
Error
Test Observation
Weighted Voting
Final Prediction
P

Leslie Valiant 分离器强弱之争

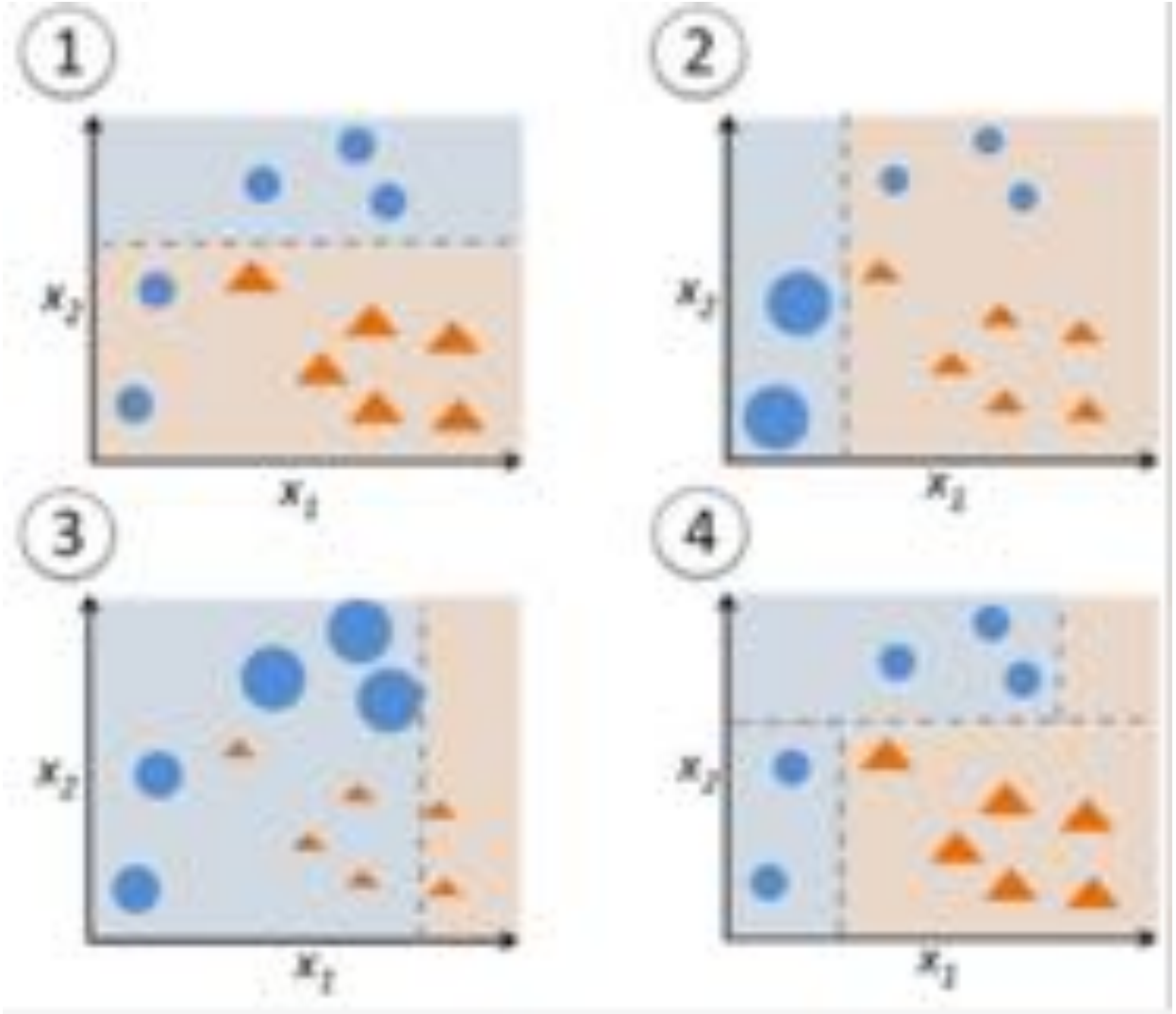
史春奇

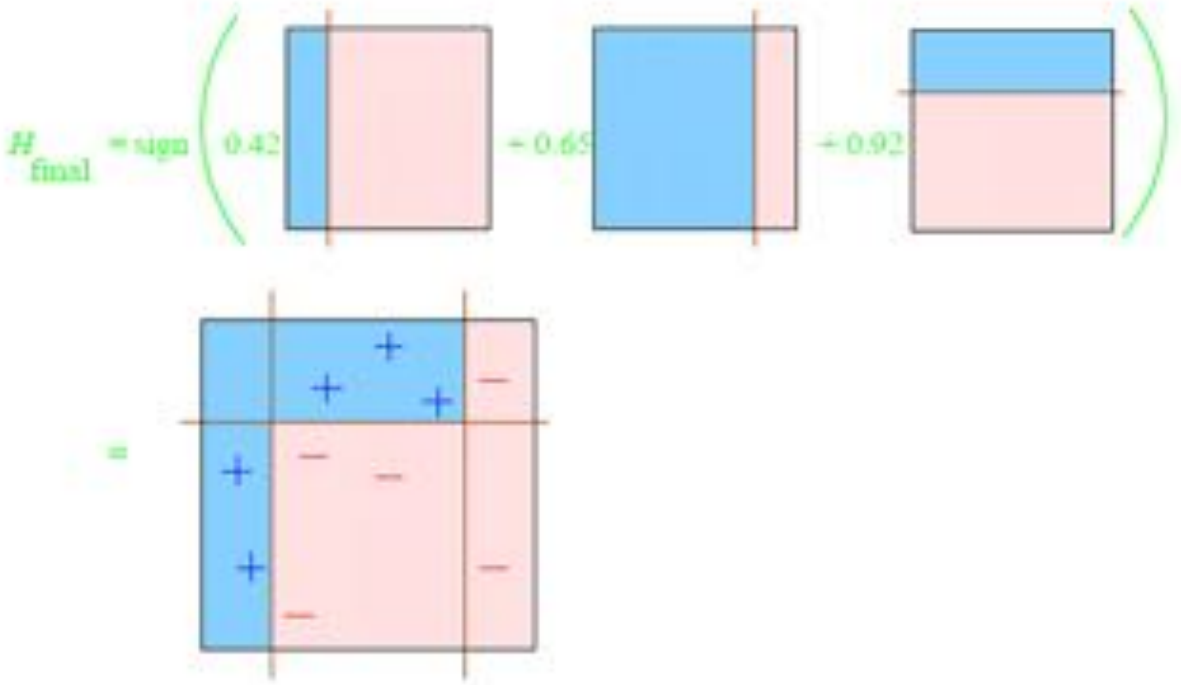
- 强分离器和弱分类器等价么？



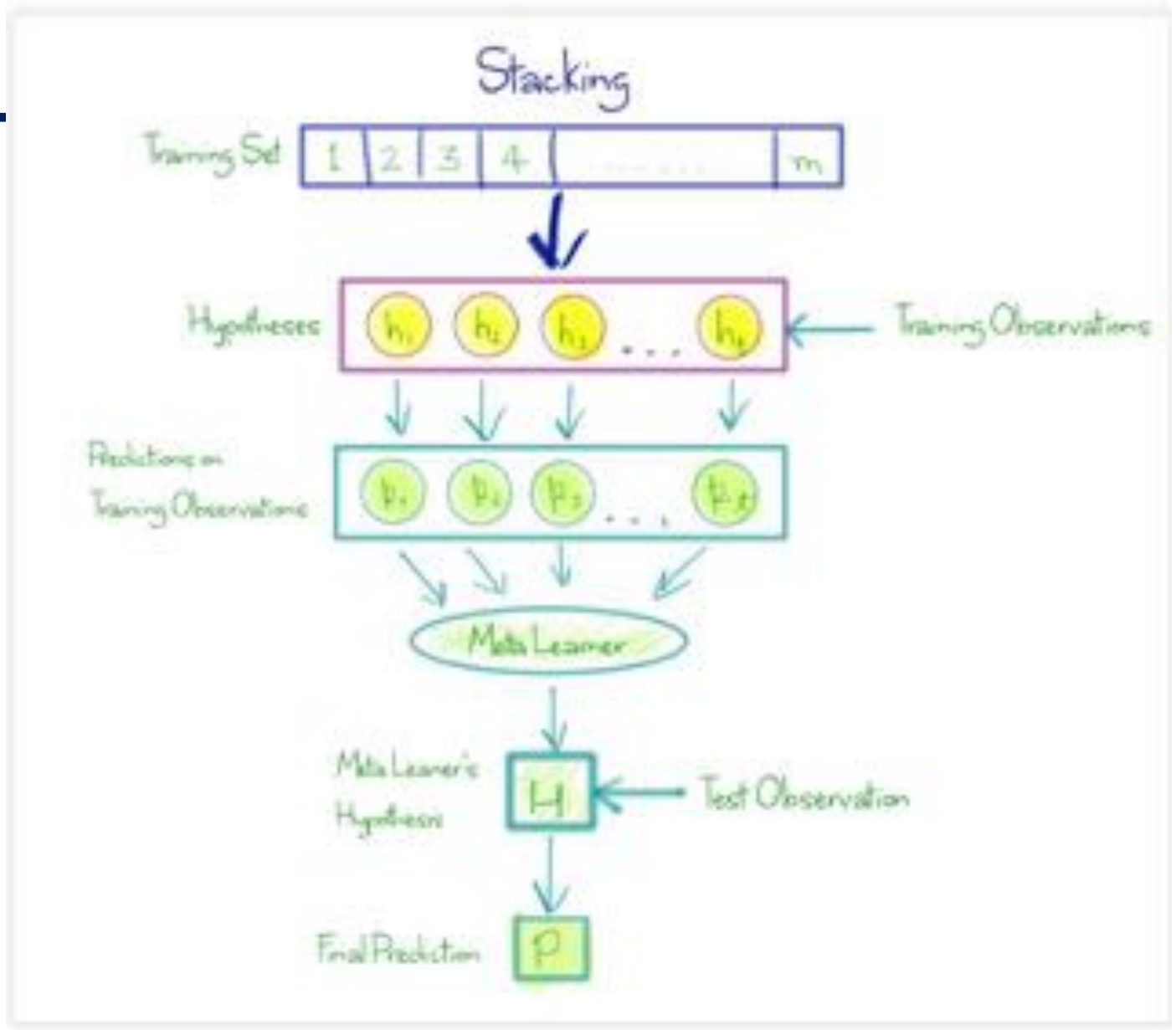


Adaboost





Stacking


Stacking
Training Set
1
2
3
4
m
Hypotheses
Training Observations
Predictions on
Training Observations
Meta Learner
Meta Learner's
Hypothesis
H
Test Observation
Final Prediction
P

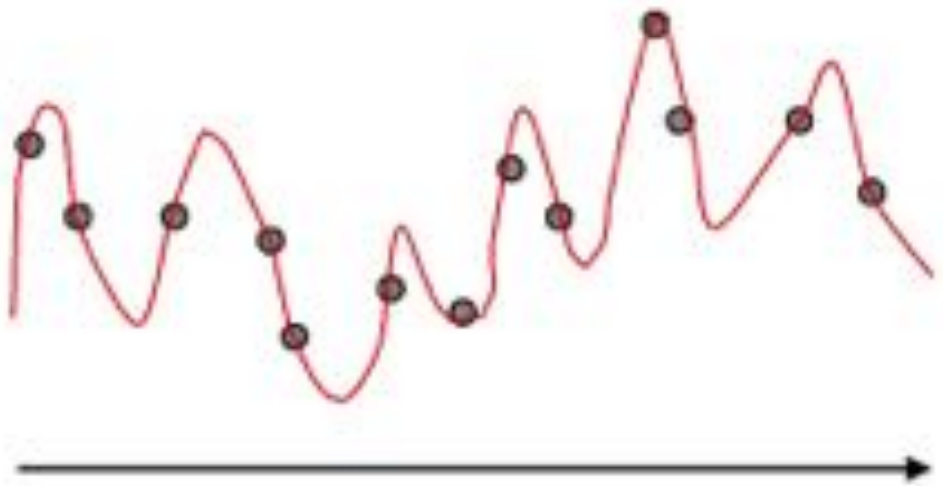
集成学习

- Bagging
  - Random Forest
  - Gradient Boosting
- **Boosting**
  - **Adaboost**
  - **Gradient Boosting**
- Stacking
  - Blending

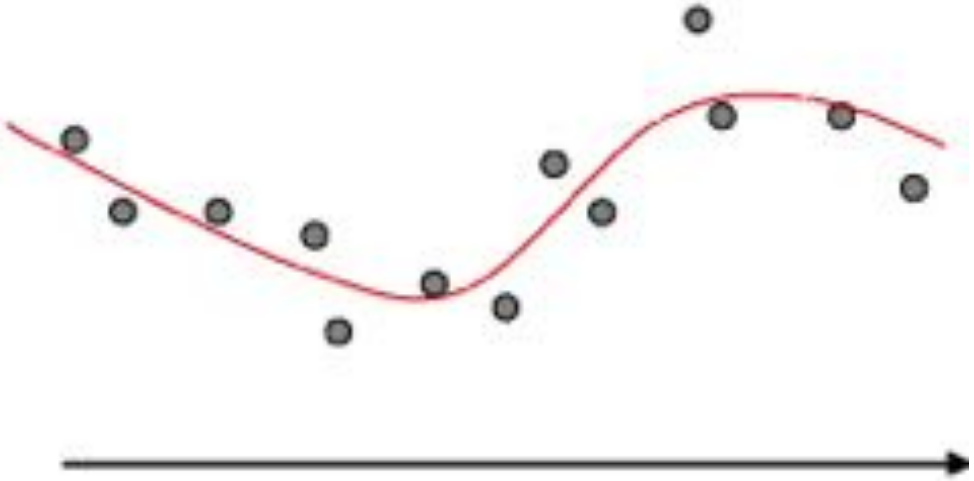
	Target	Data	Classifier	Aggregation
Bagging	mainly variance	replacement resampling (bootstrap) based randomization (parallel)	homogeneous uncorrelated (sub)-learner	voting for classification
				averaging for regression
Boosting	mainly bias	stepwise misclassification based decorrelation (sequential)	homogeneous weak learner	weighted majority voting
Stacking	both bias and variance	various	heterogeneous strong learner	meta-learner

正则化 – 回归和分类

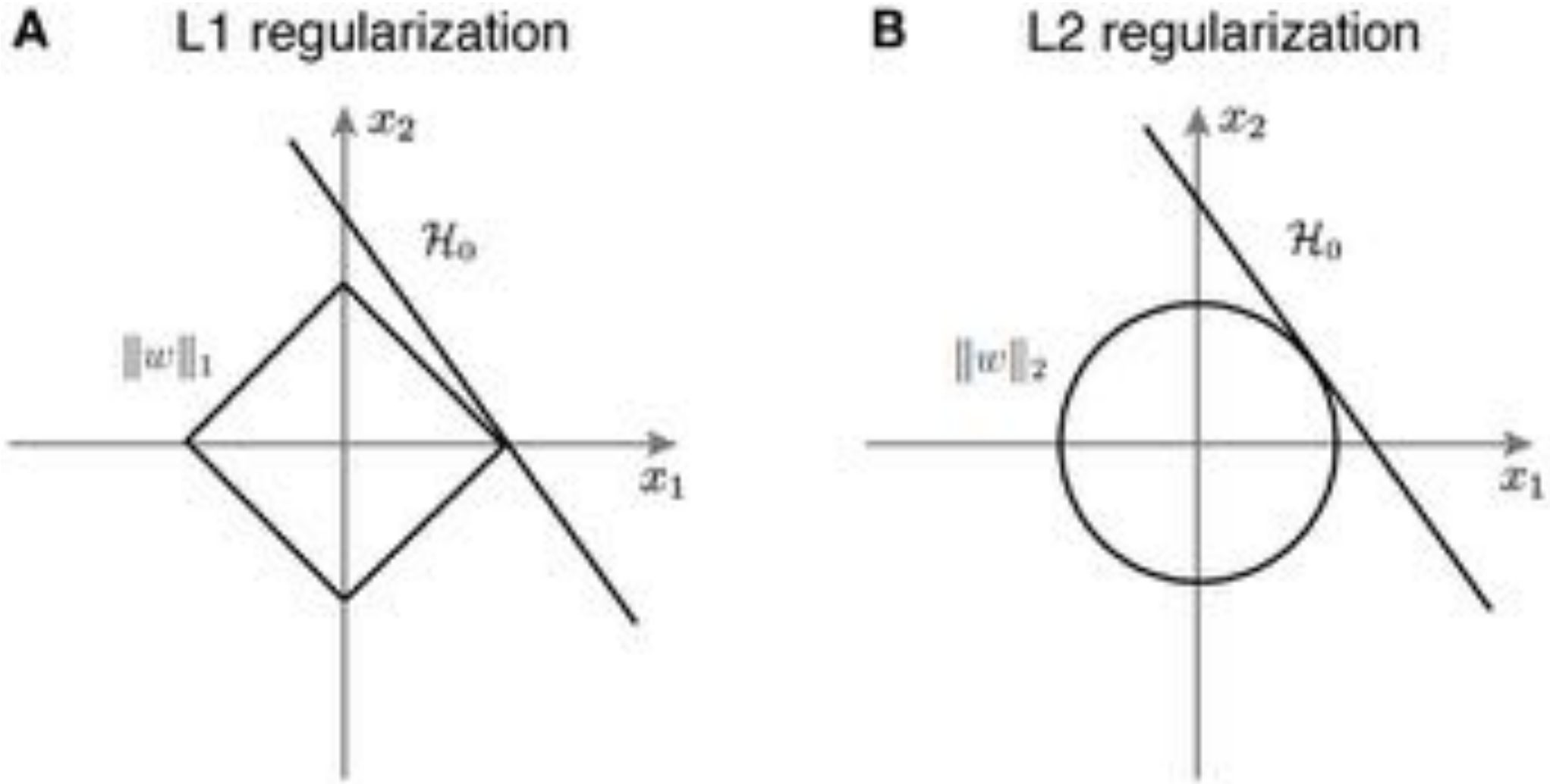
$$\min_f |Y_i - f(X_i)|^2$$



$$\min_{f \in H} \sum_{i=1}^{n} |Y_i - f(X_i)|^2 + \lambda \|f\|_H^2$$

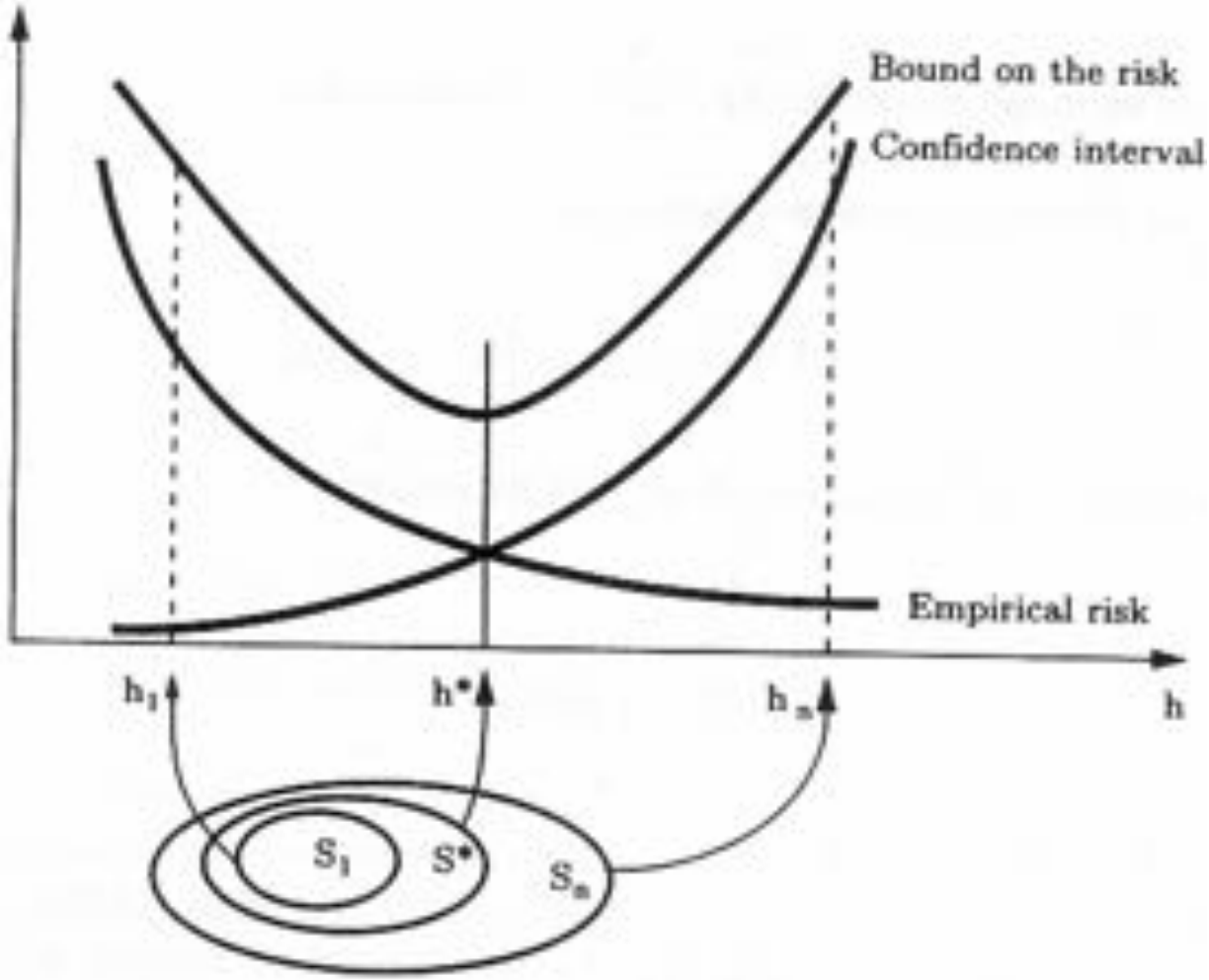


正则化 与 参数先验


A
L1 regularization
x_2
$\mathcal{H}_0$
$\|w\|_1$
x_1
B
L2 regularization
x_2
$\mathcal{H}_0$
$\|w\|_2$
x_1

结构风险最小 Structural Risk Minimization

史春奇

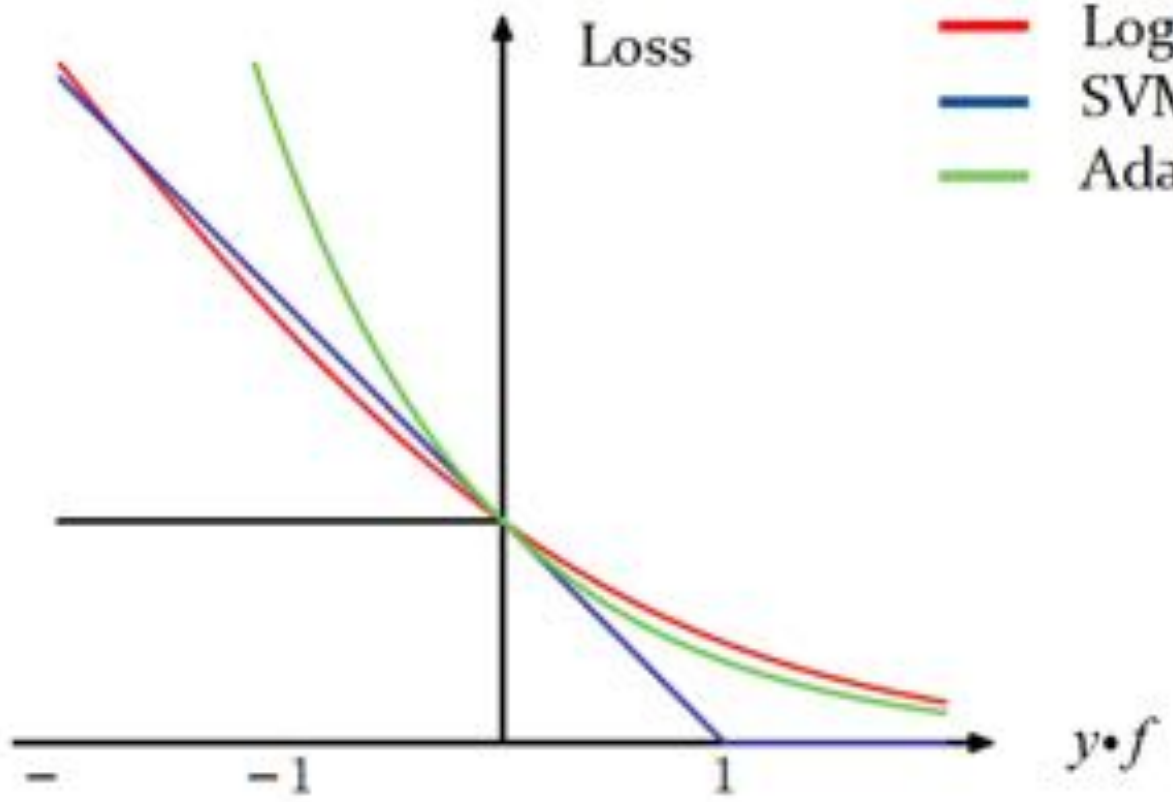


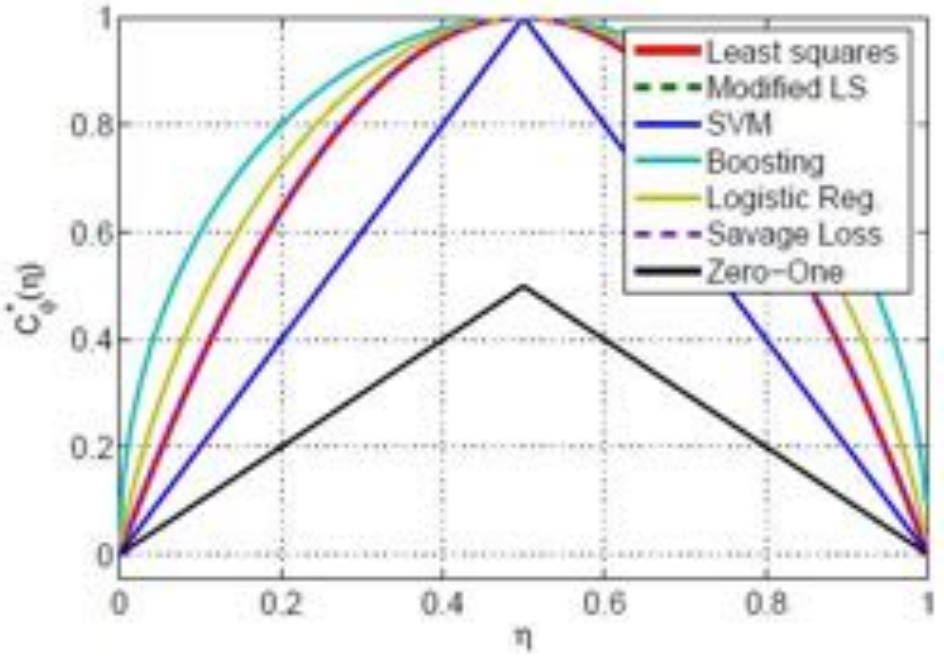
$$l_{total}(\vec{w},\vec{x}) = l_{model}(\vec{w},\vec{x}) + l_{reg}(\vec{w})$$

L2 Regularization: $l_{reg}(\vec{w}) = \lambda \sum_{i=1}^{N} w_i^2$

L1 Regularization: $l_{reg}(\vec{w}) = \lambda \sum_{i=1}^{N} |w_i|$

分类算法 在 正则化下统一





ADVANCED?

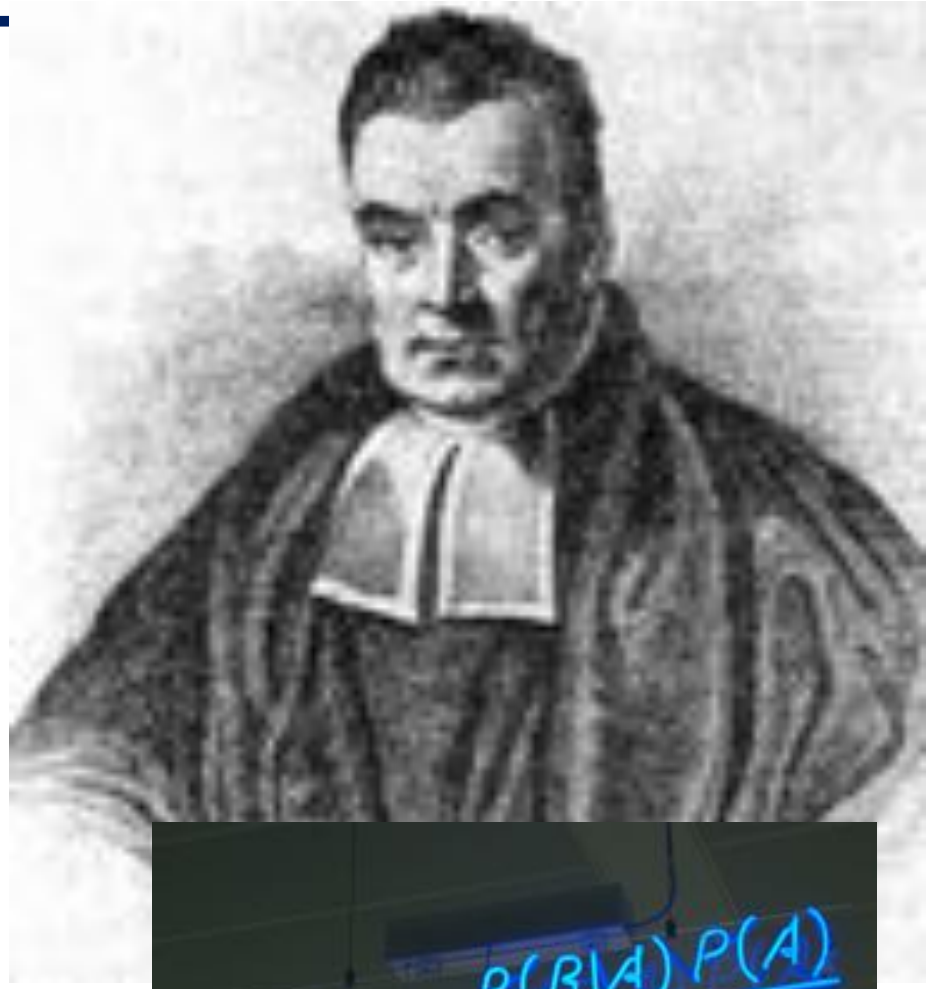
- 统计学习高级阶段？

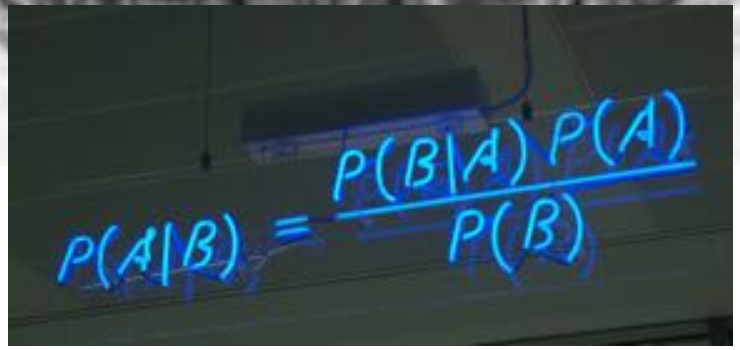
贝叶斯网络和概率图模型

托马斯·贝叶斯 Thomas Bayes

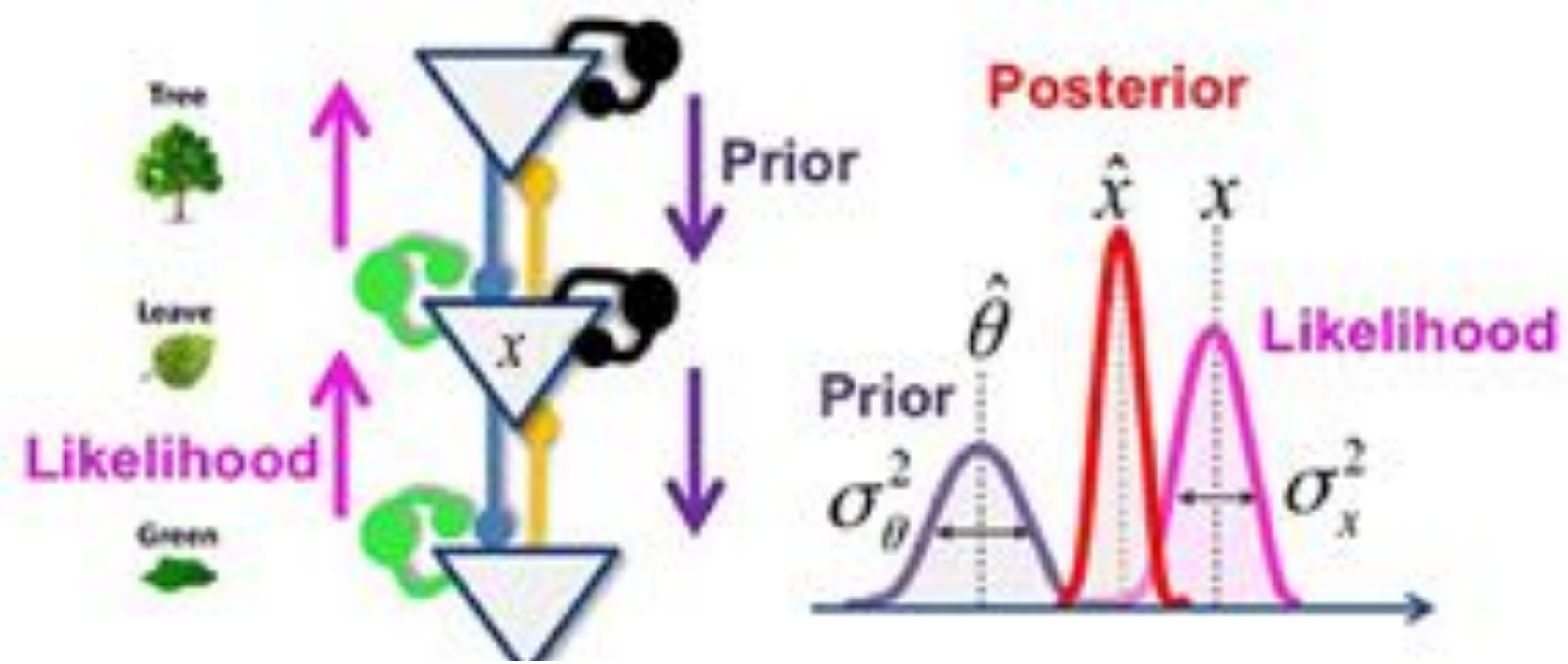
- 1702年 - 1761年
- 英国数学家
- 贝叶斯公式

$$P(A \mid B) = \frac{P(B \mid A)\,P(A)}{P(B)}$$





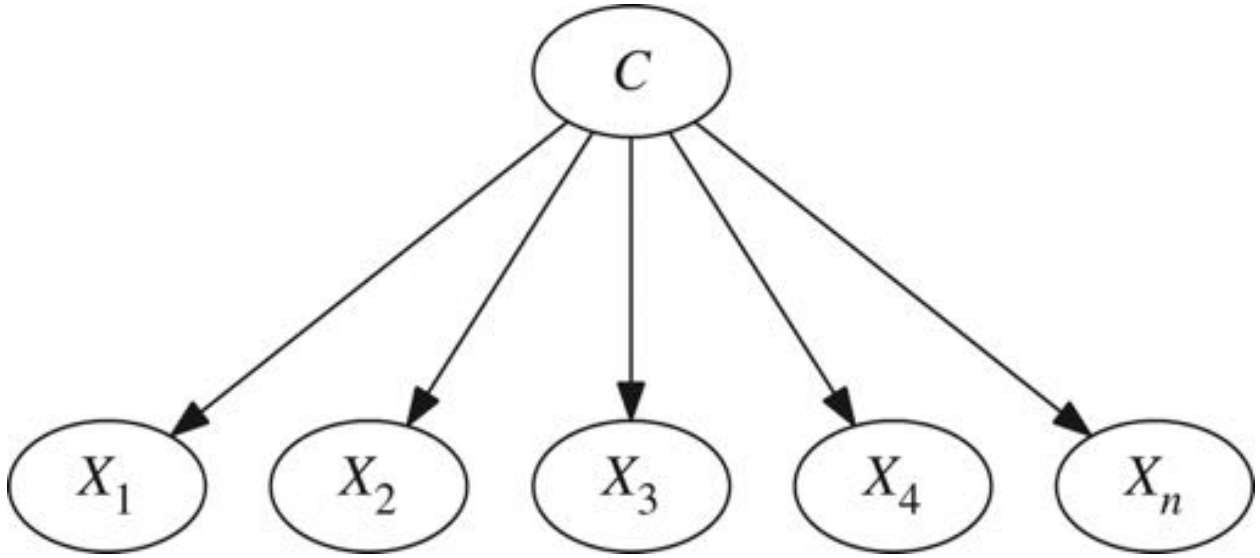
贝叶斯公式 – 数据解读 – 先验



$$\text{Posterior } \alpha \text{ Prior } \times \text{ Likelihood}$$

$$P(\theta \mid \text{Data}) = \frac{P(\theta)\ P(\text{Data} \mid \theta)}{P(\text{Data})}$$

贝叶斯分类



$$\text{Posterior} = \frac{\text{Likelihood}^* \text{ Prior}}{\text{Evidence}}$$

$$P(C_j \mid A_1, A_2, \ldots, A_n) = \frac{\left(\prod_{i=1}^{n} P(A_i \mid C_j)\right) P(C_j)}{P(A_1, A_2, \ldots, A_n)}$$

Naïve Bayes 分类器 举例

$P(x \mid c) = P(\text{Sunny} \mid \text{Yes}) = 3/9 = 0.33$

Frequency Table		Play Golf	
		Yes	No
Outlook	Sunny	3	2
	Overcast	4	0
	Rainy	2	3

Likelihood Table		Play Golf		
		Yes	No	
Outlook	Sunny	3/9	2/5	5/14
	Overcast	4/9	0/5	4/14
	Rainy	2/9	3/5	5/14
		9/14	5/14	

$P(x) = P(\text{Sunny}) = 5/14 = 0.36$

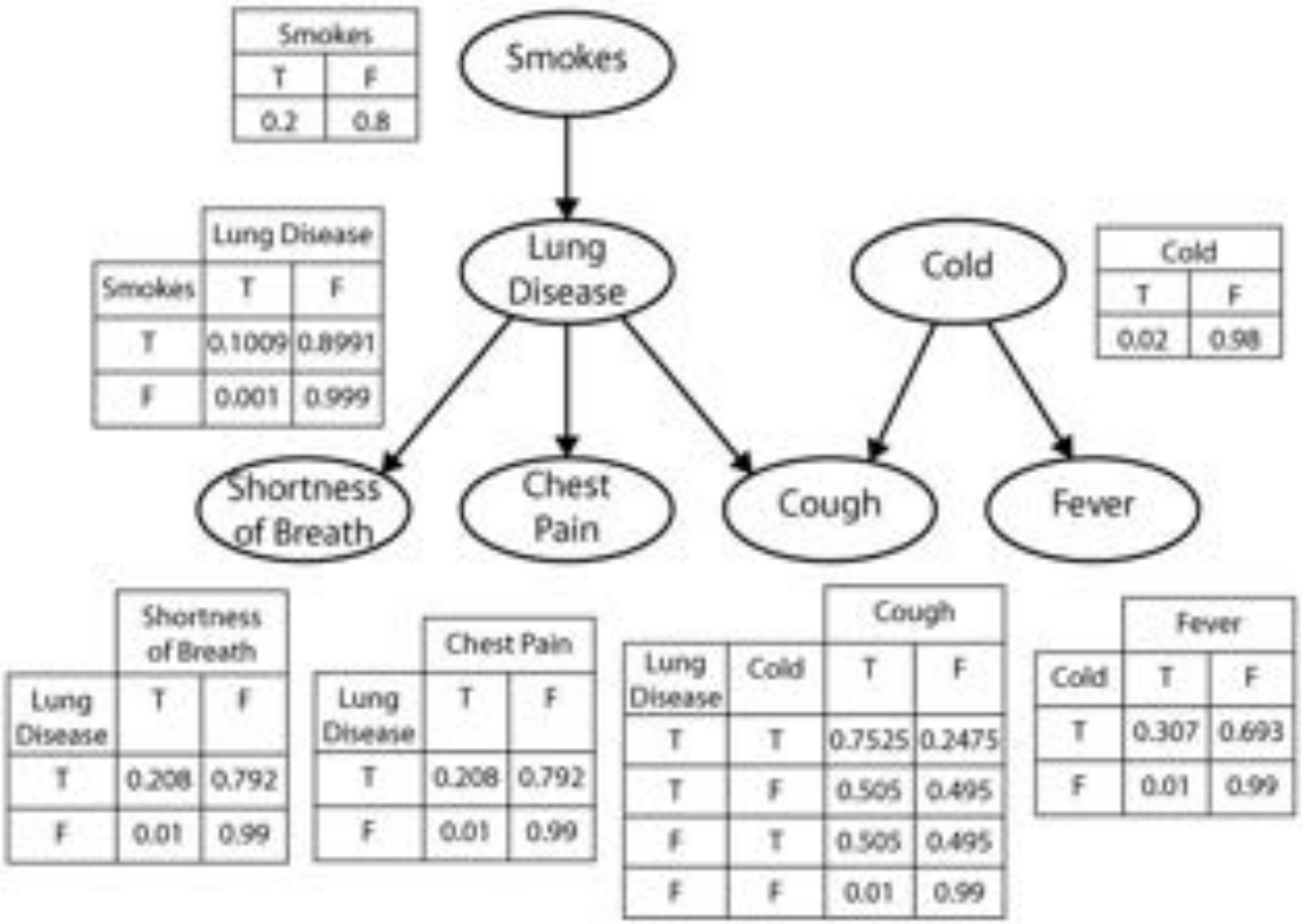
$P(c) = P(\text{Yes}) = 9/14 = 0.64$

Posterior Probability:

$P(c \mid x) = P(\text{Yes} \mid \text{Sunny}) = 0.33 \times 0.64 \div 0.36 = 0.60$

贝叶斯推理

史春奇



Smokes	
T	F
0.2	0.8

Smokes	Lung Disease	
	T	F
T	0.1009	0.8991
F	0.001	0.999

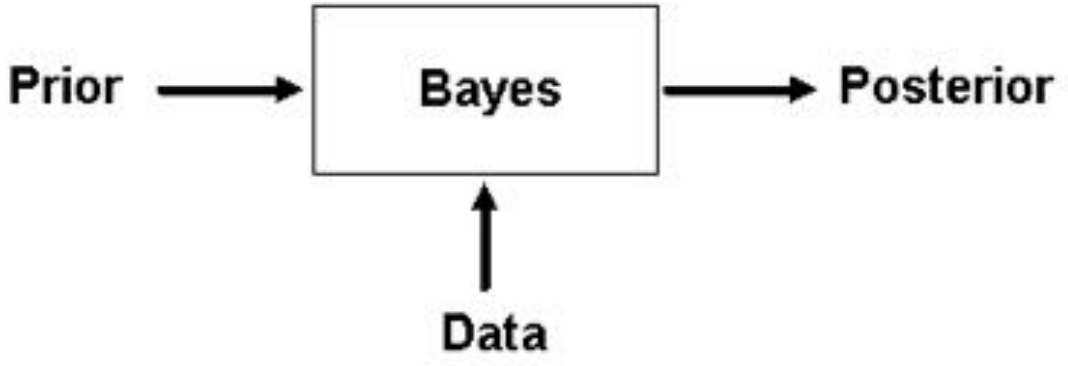
Cold	
T	F
0.02	0.98

Lung Disease	Shortness of Breath	
	T	F
T	0.208	0.792
F	0.01	0.99

Lung Disease	Chest Pain	
	T	F
T	0.208	0.792
F	0.01	0.99

Lung Disease	Cold	Cough	
		T	F
T	T	0.7525	0.2475
T	F	0.505	0.495
F	T	0.505	0.495
F	F	0.01	0.99

Cold	Fever	
	T	F
T	0.307	0.693
F	0.01	0.99



Posterior probability

Likelihood

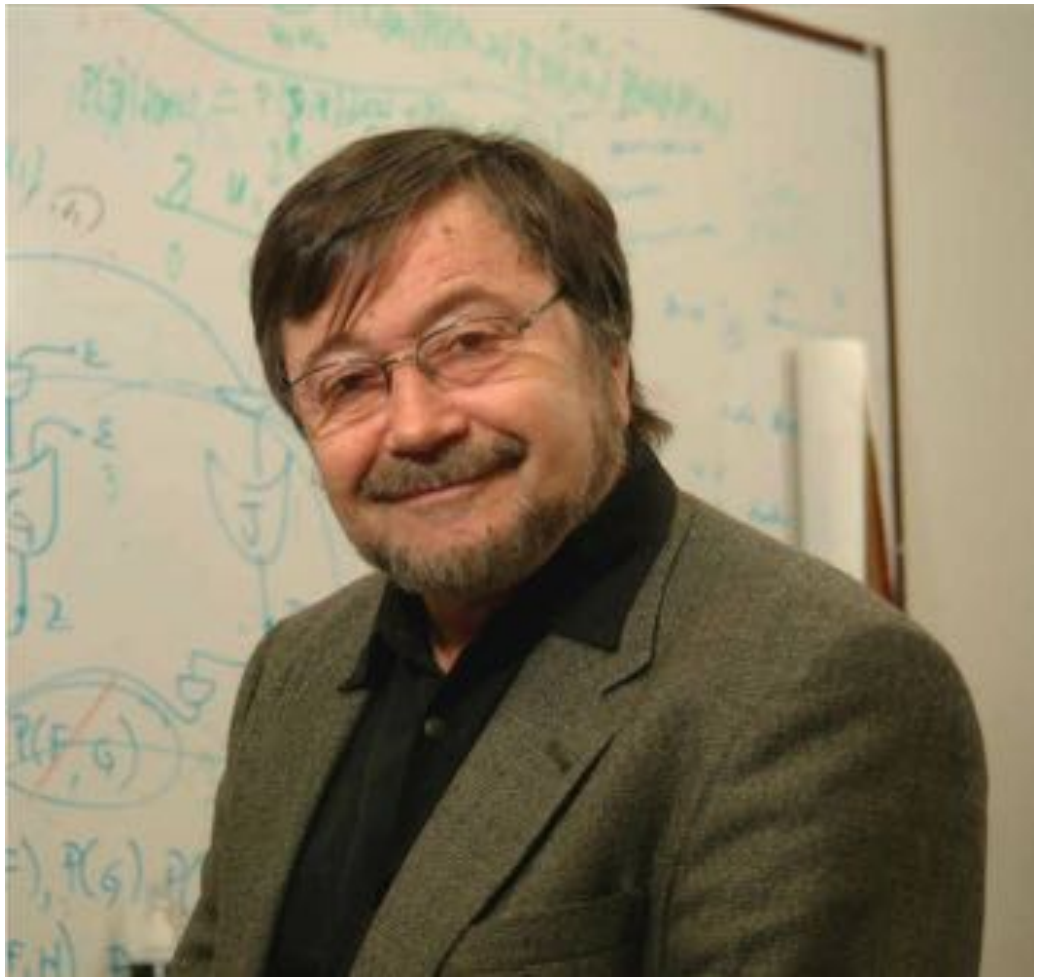
Prior probability

$$p(A|B) = \frac{p(B|A)\,p(A)}{p(B)}$$

朱迪亚·珀尔 Judea Pearl

- 1936 出生
- Rutgers University 罗格斯大学 硕士
- 纽约大学坦登工程学院 New York University Tandon School of Engineering 博士

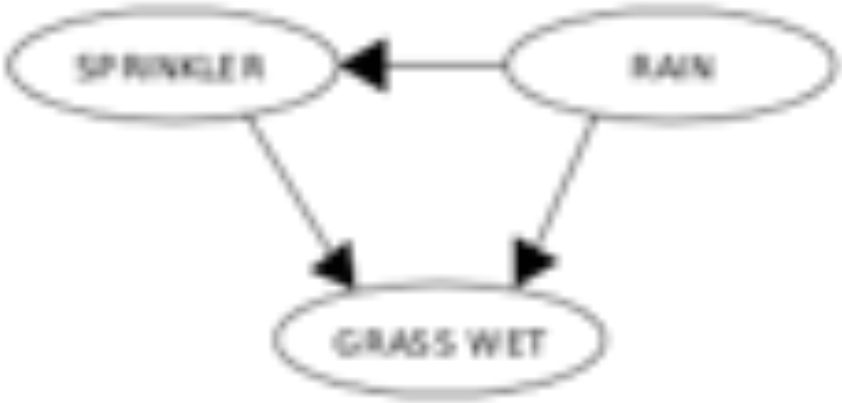
Splines



- UCLA 认知实验室主任, 统计和计算机系教授
- 贝叶斯网络提出者
- 图灵奖获得者

贝叶斯网络

SPRINKLER		
RAIN	T	F
F	0.4	0.6
T	0.01	0.99



RAIN	
T	F
0.2	0.8

		GRASS WET	
SPRINKLER	RAIN	T	F
F	F	0.0	1.0
F	T	0.8	0.2
T	F	0.9	0.1
T	T	0.99	0.01

马克尔·乔丹 Michael I. Jordan

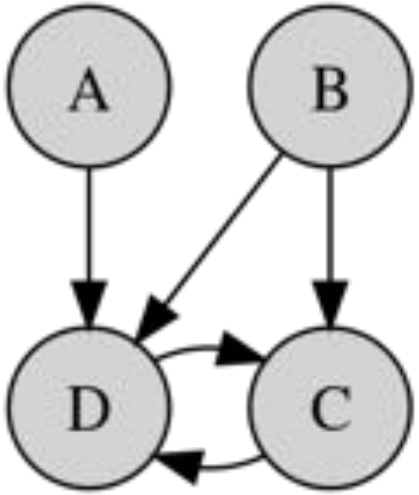
- 出生1956年
- Arizona State University硕士
- UCSD博士

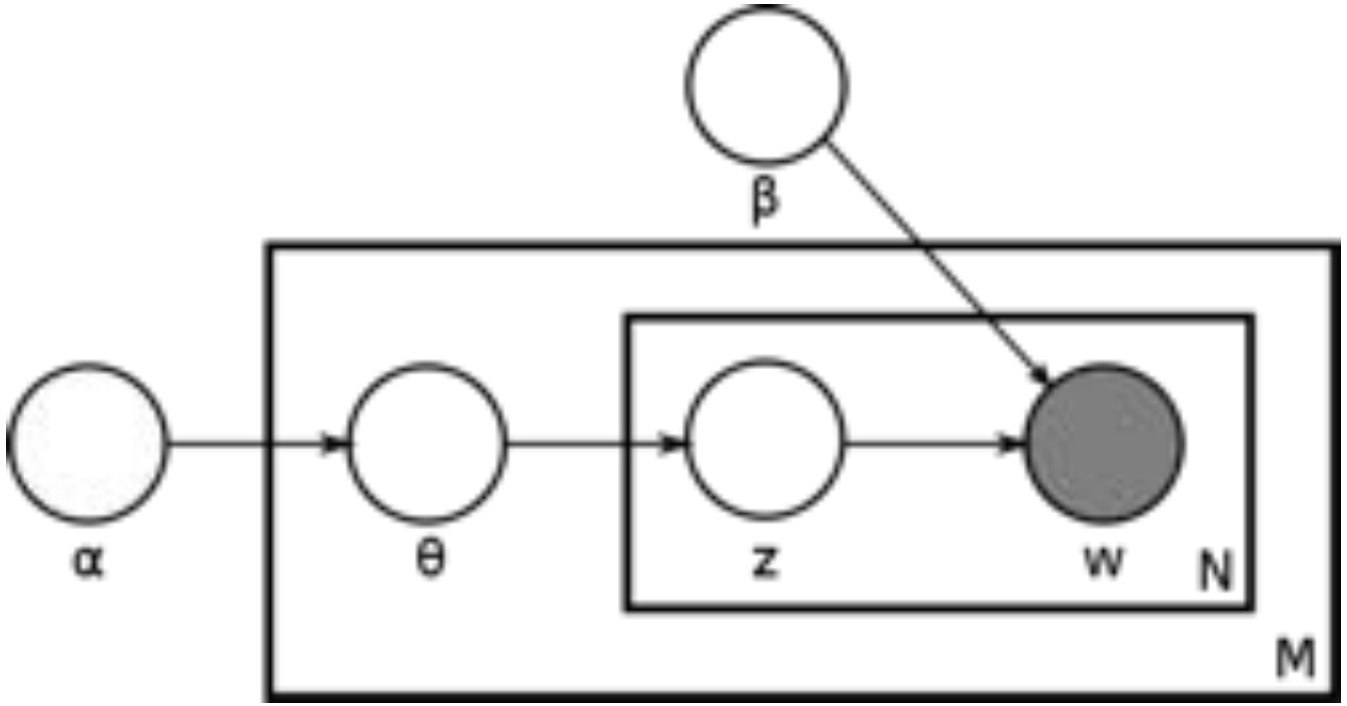
- 伯克利教授
  - Breiman从MIT引荐到伯克利
  - 博士弟子中名人超多
    - 吴恩达 Andrew Ng， 张志华， 邢波

- 概率图模型的主要贡献者

概率图模型

- Probabilistic graphical model
  - Markov random field
  - HMM, PLA, LDA, CRF





CHINESE?

- 国内代表人物？

南北二枝花（志华）

- 南京大学， 周志华 (集成学习 等等)
- 北京大学， 张志华 (图模型 等等)

还有很多很多！

MORE

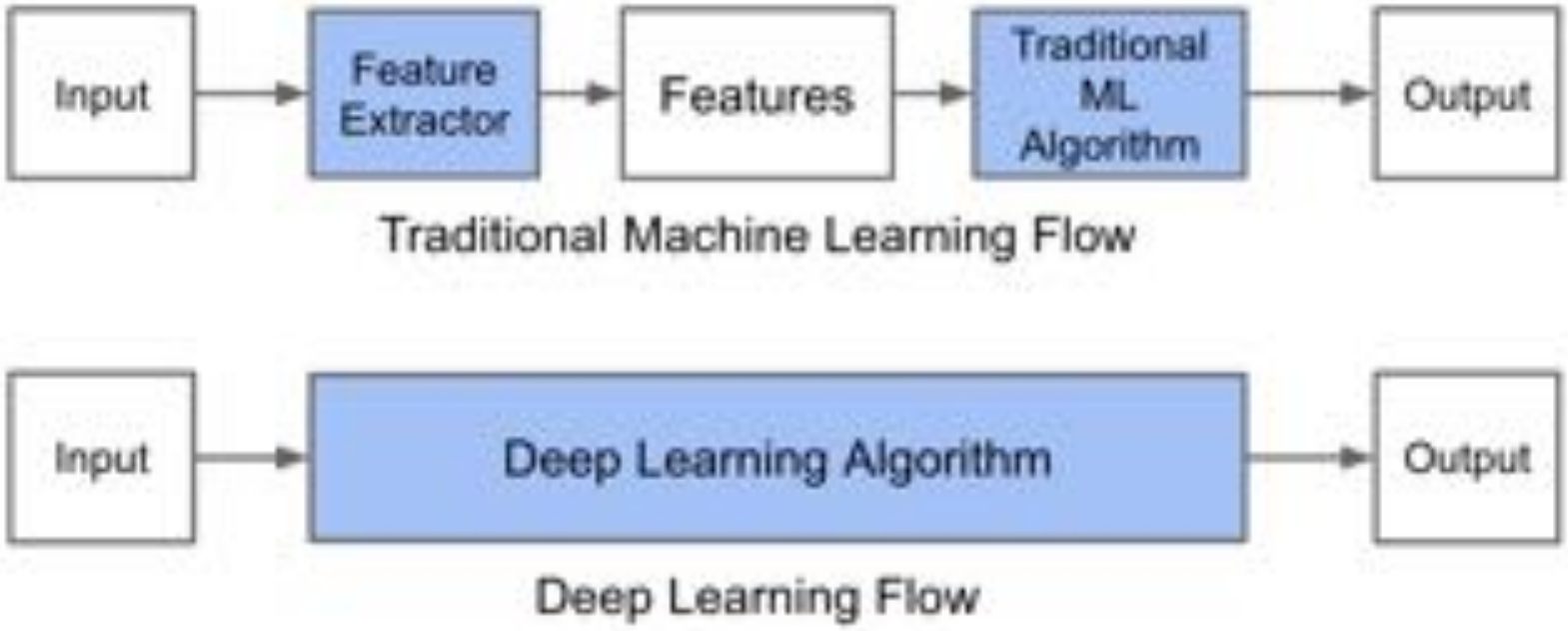
MORE

WHY DEEP LEARNING?

- 为什么深度学习会崛起？

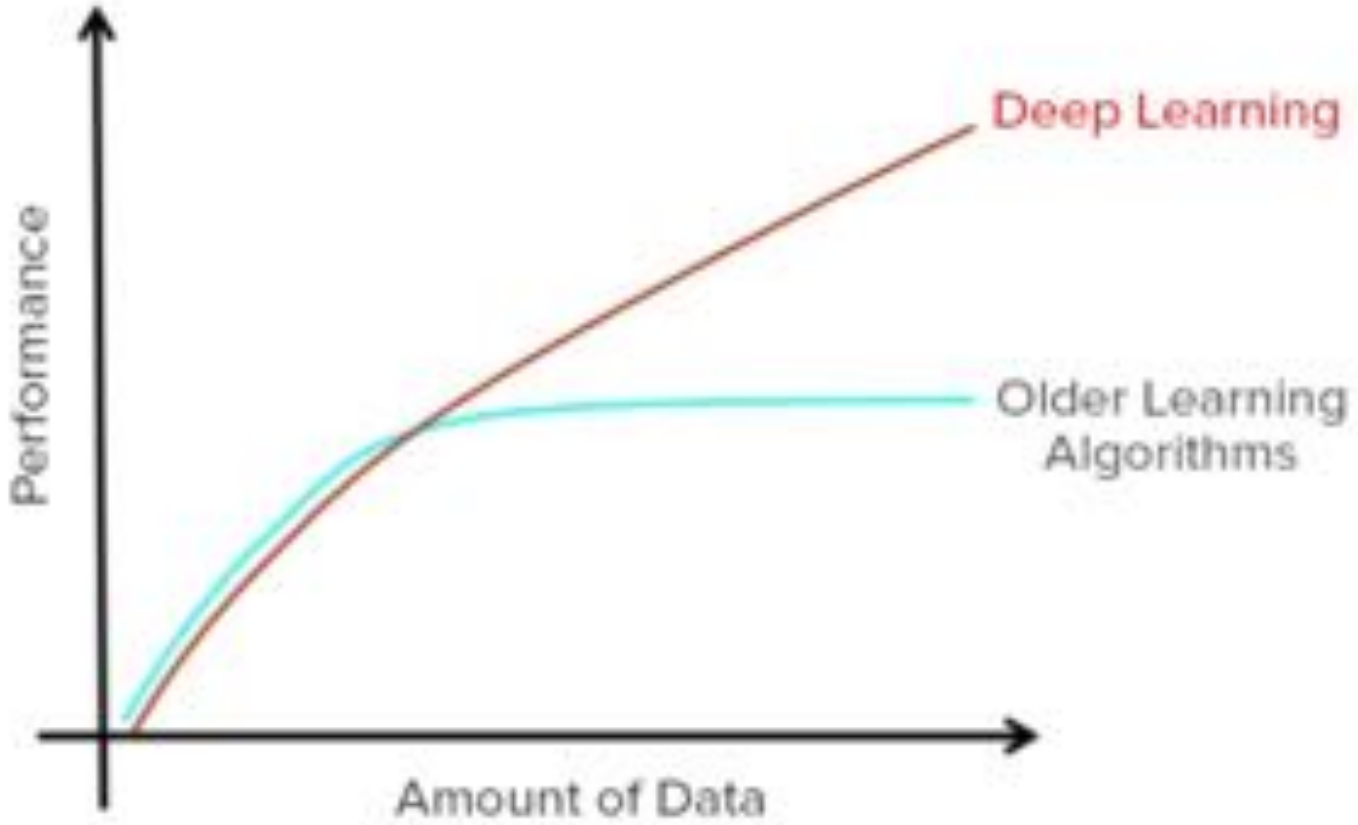
统计学习局限性

- 强大的特征表达能力



深度模型强大的拟合能力

- 效果对比



Geoffrey E. Hinton 辛顿老仙， 法力无边

- 出生1947年
- University of Edinburgh博士
- University of Toronto 教授
  - Google
  - Carnegie Mellon University
  - University College London
- BP算法 Backpropagation
- 玻尔兹曼机 Boltzmann machine
- 图模型 Probabilistic graphical model
- 深度学习 Deep learning
- 博士后超级厉害！

和Vapnik可以一拼的引用量

Geoffrey Hinton

Emeritus Professor of Computer Science, University of Toronto & Engineering Fellow, Google Inc

machine learning, neural networks, artificial intelligence, cognitive science, computer science

Verified email at cs.toronto.edu - Homepage

Citation indices	All	Since 2012
Citations	176096	82531
h-index	130	96
i10-index	308	221

Co-authors View all...

Terrence Sejnowski, James L. McClelland, Abdel-rahman Mohamed

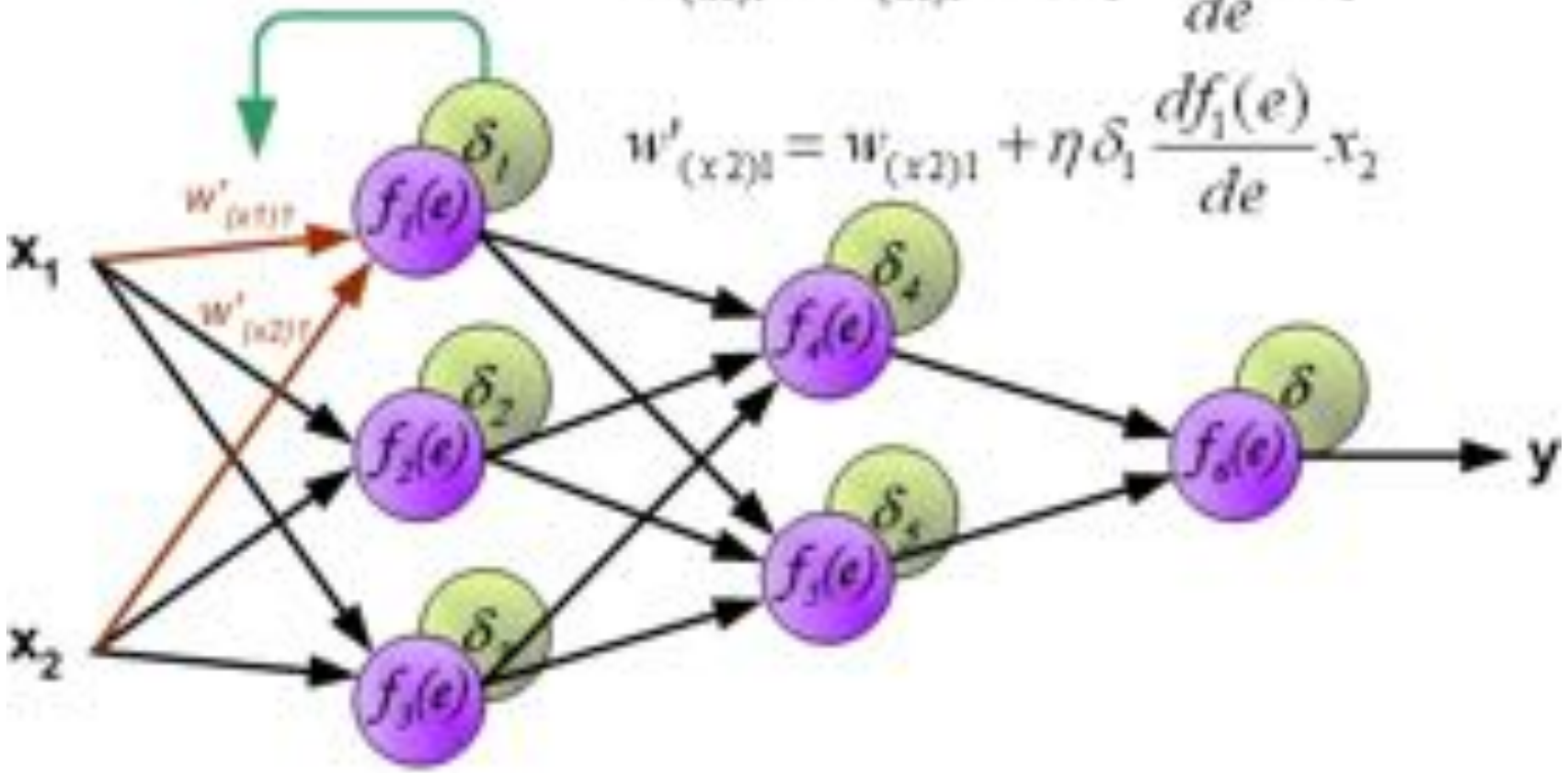
Title 1–20	Cited by	Year
Learning internal representations by error-propagation DE Rumelhart, GE Hinton, RJ Williams Parallel Distributed Processing: Explorations in the Microstructure of ...	23308	1986
Learning internal representations by error propagation DE Rumelhart, GE Hinton, RJ Williams CALIFORNIA UNIV SAN DIEGO LA JOLLA INST FOR	~~23016~~	1985
Parallel distributed processing DE Rumelhart, JL McClelland, PDP Research Group IEEE 1, 443-453	21747	1988
Learning representations by back-propagating errors DE Rumelhart, GE Hinton, RJ Williams Nature 323, 533-536	12575	1986

The Deep Learning Saga

Backpropagation

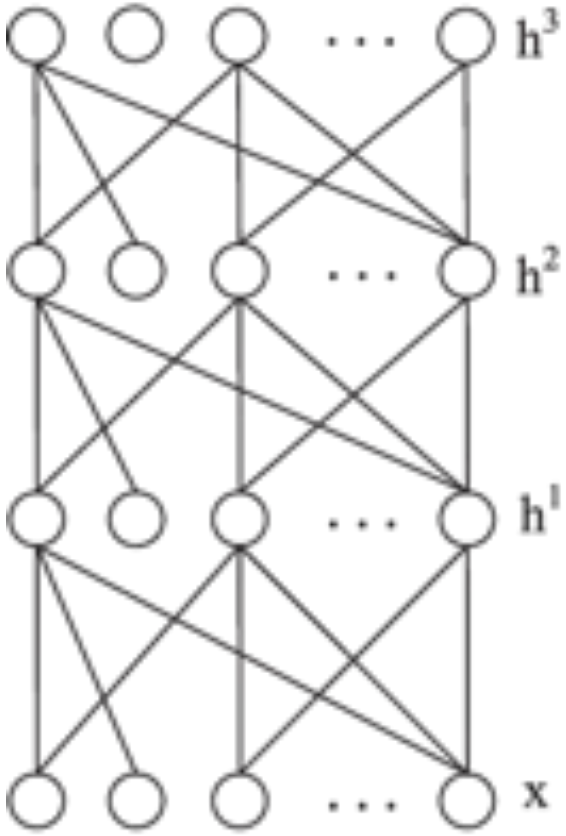
$$w'_{(x1)1} = w_{(x1)1} + \eta\delta_1 \frac{df_1(e)}{de} x_1$$

$$w'_{(x2)1} = w_{(x2)1} + \eta\delta_1 \frac{df_1(e)}{de} x_2$$

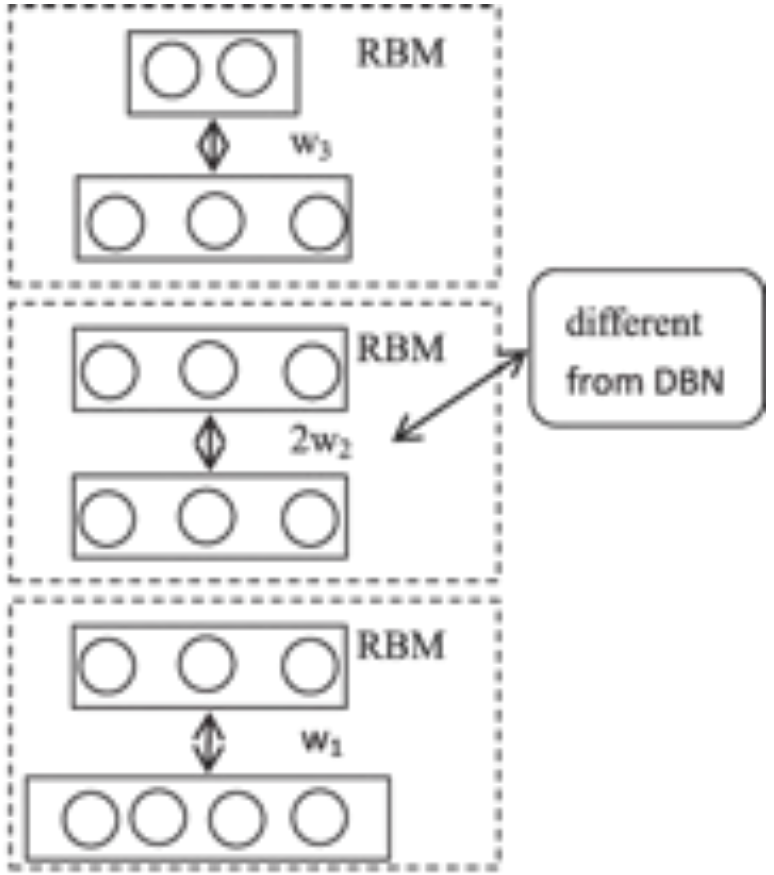


DBM

- Deep Boltzmann machine

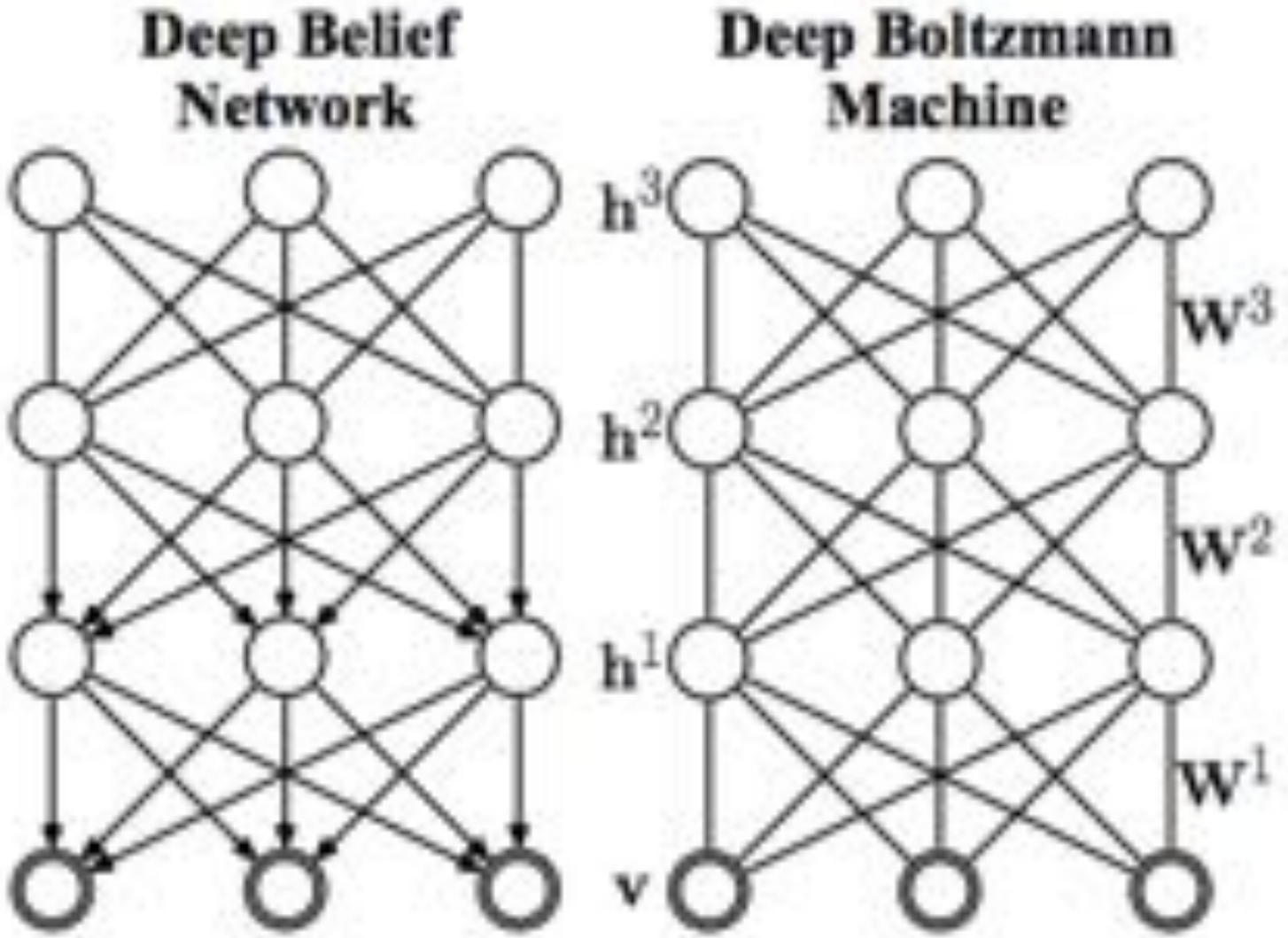


Deep Boltzmann Machine



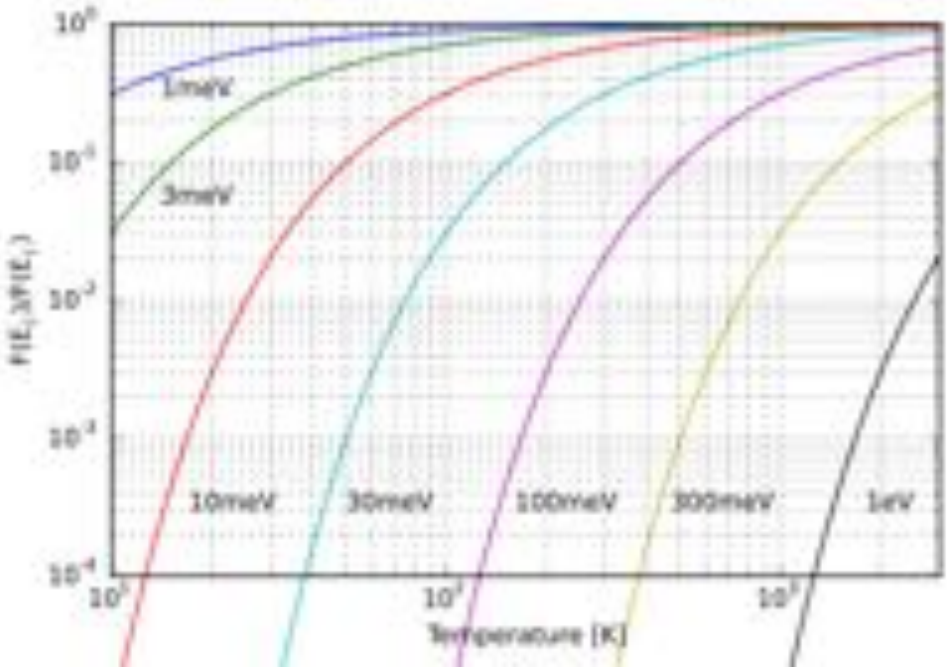
Pre-training of DBM

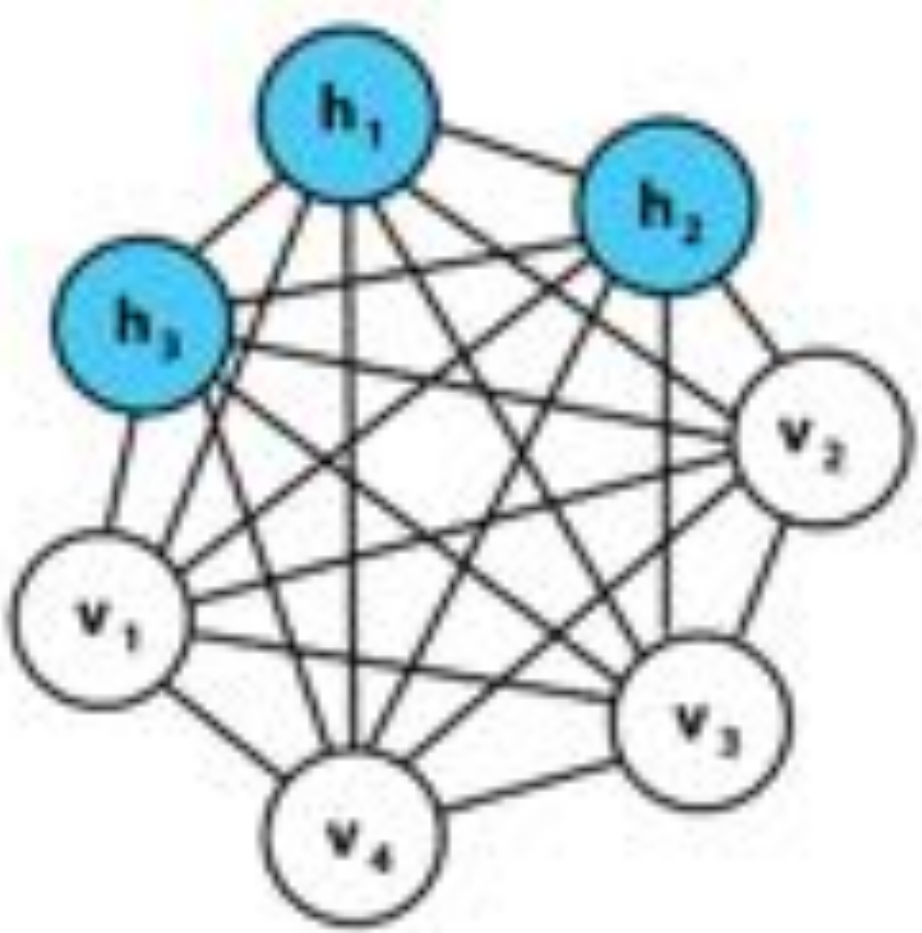
DBM vs DBN


Deep Belief Network
Deep Boltzmann Machine
h3
h2
h1
v
W3
W2
W1

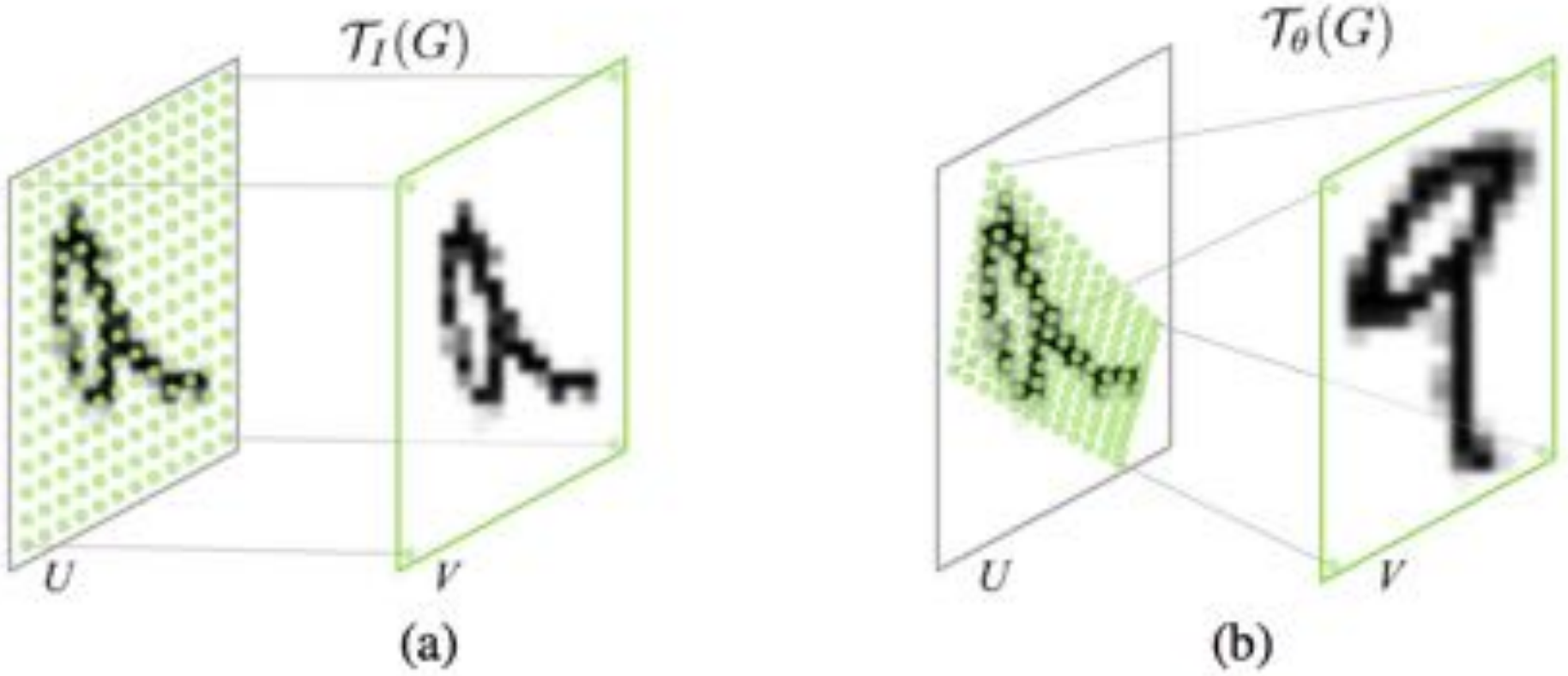
Boltzmann Machine

- Boltzmann distribution + Hopfield Net





Spatial Transformation



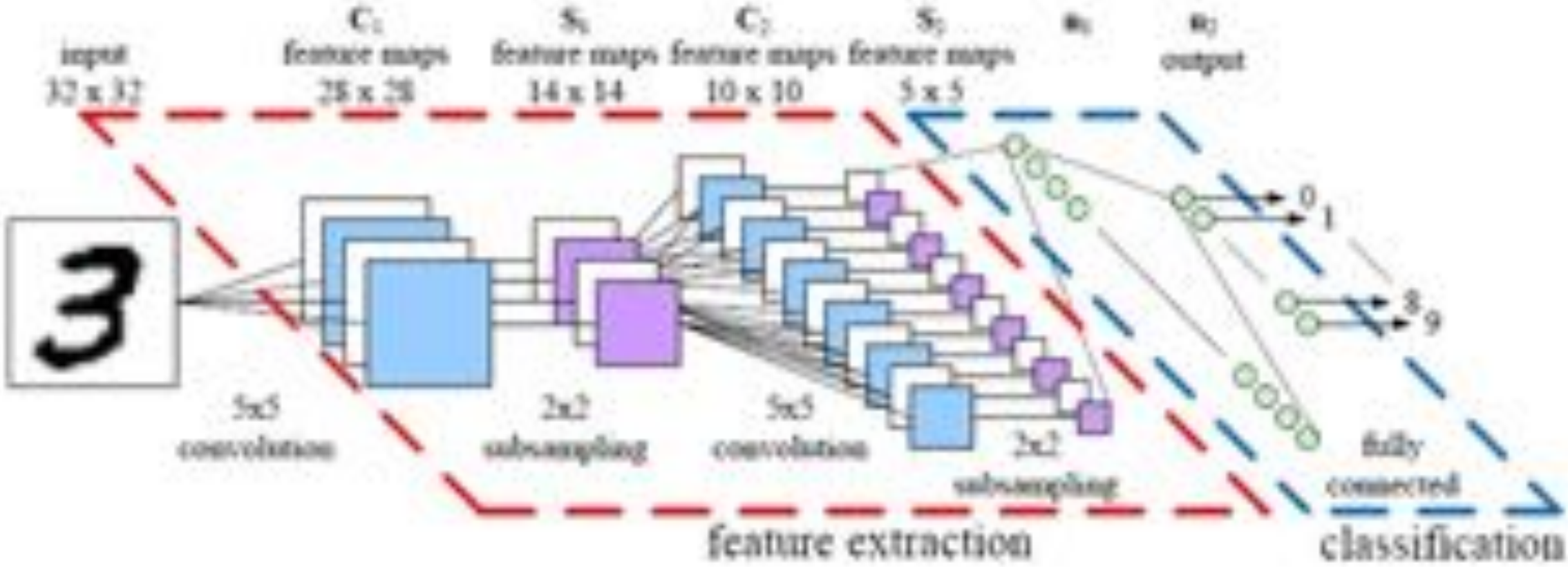
Yann LeCun

- 出生1960年
- Pierre and Marie Curie University 巴黎第六大学博士
- 辛顿的博士后

- Facebook的研究员

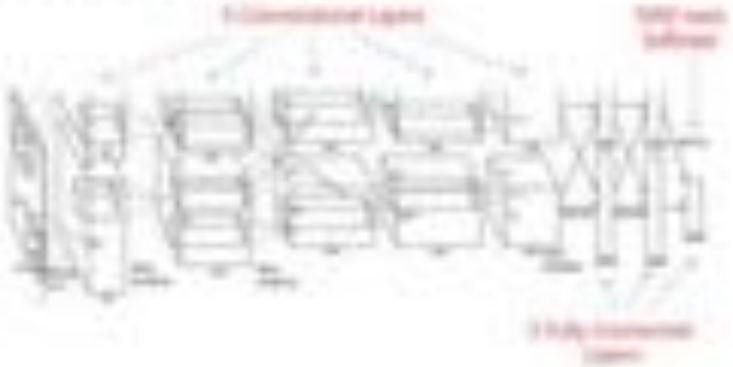
卷积神经网络 CNN

- convolutional neural network



卷积神经网络 架构变换

AlexNet



VGG

GoogLeNet

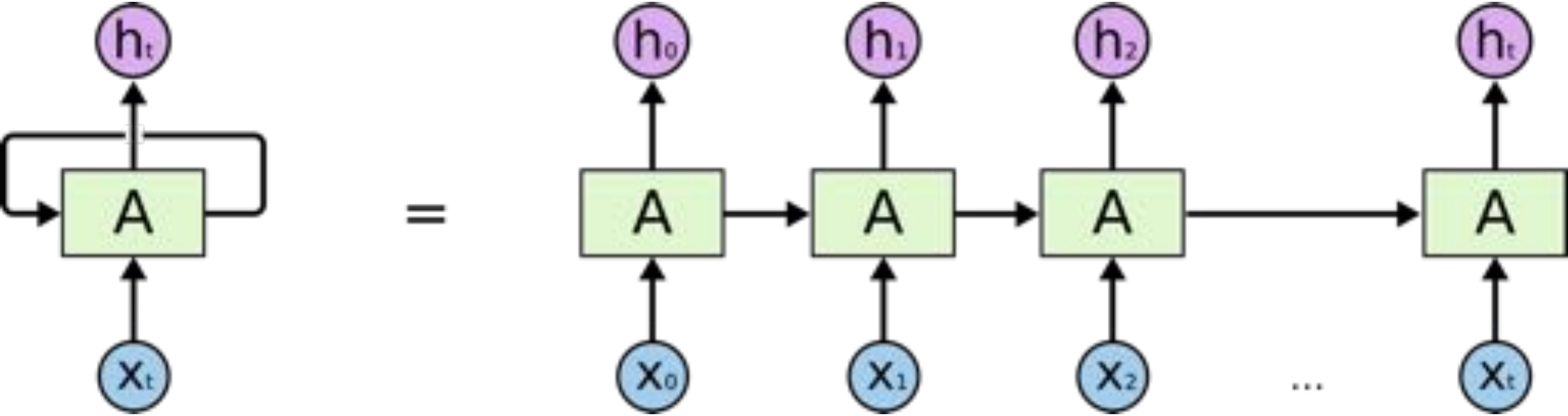
ResNet

Jürgen Schmidhuber

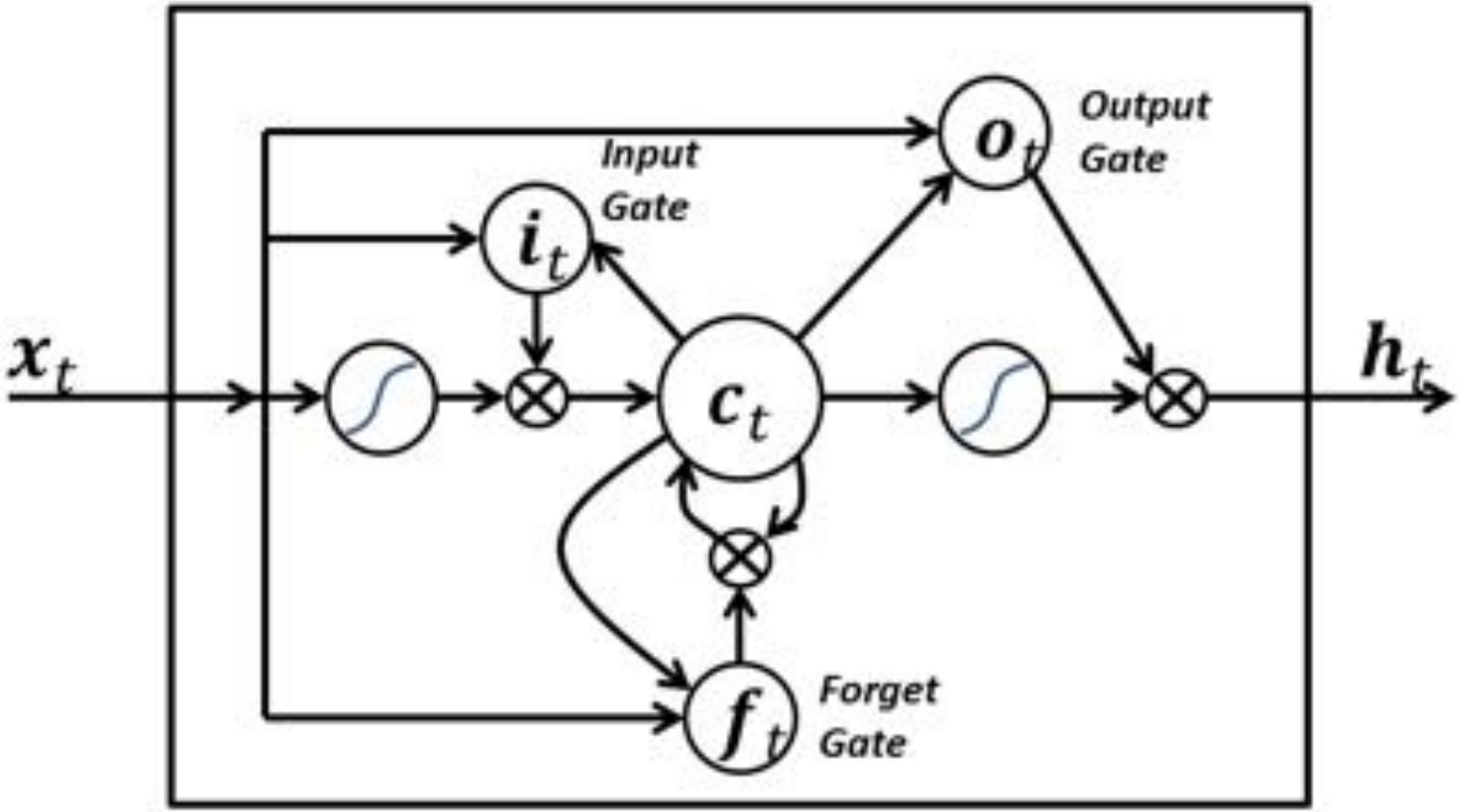
- 出生1963年
- Technical University of Munich
  慕尼黑工业大学博士

- Recurrent Neural Network, RNN

Recurrent Neural Network



Long short-term memory

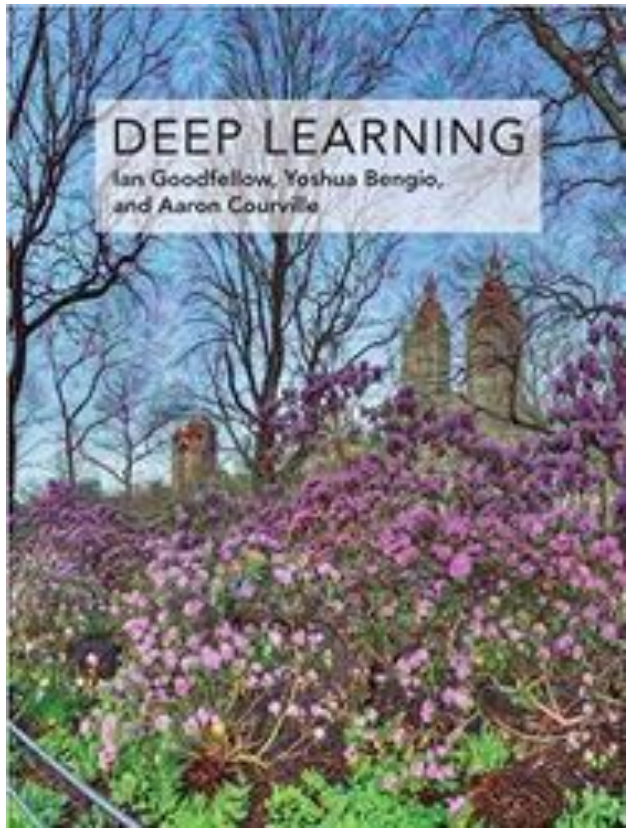


Yoshua Bengio

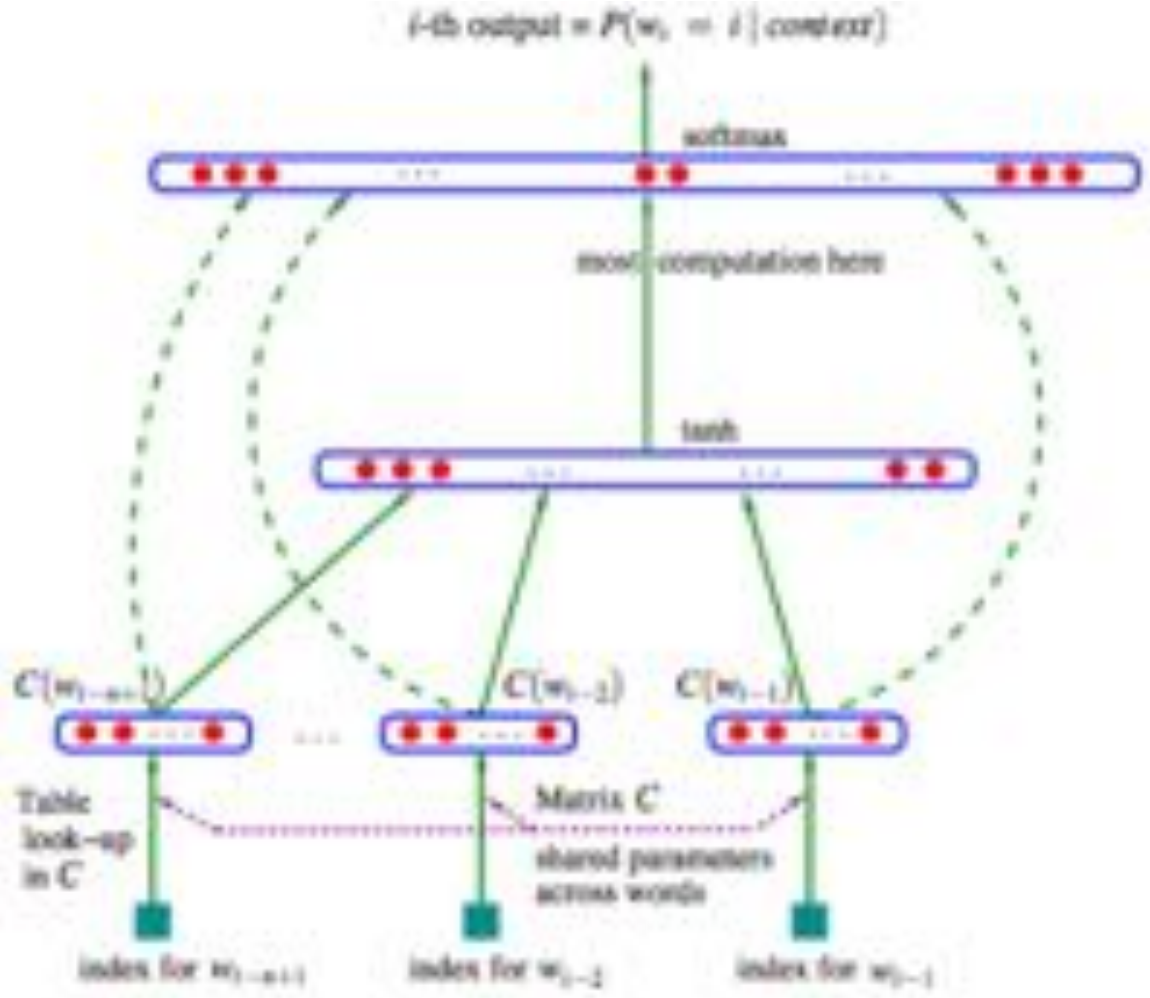
- 出生1964年
- McGill University博士
- Université de Montréal
  蒙特利尔大学教授

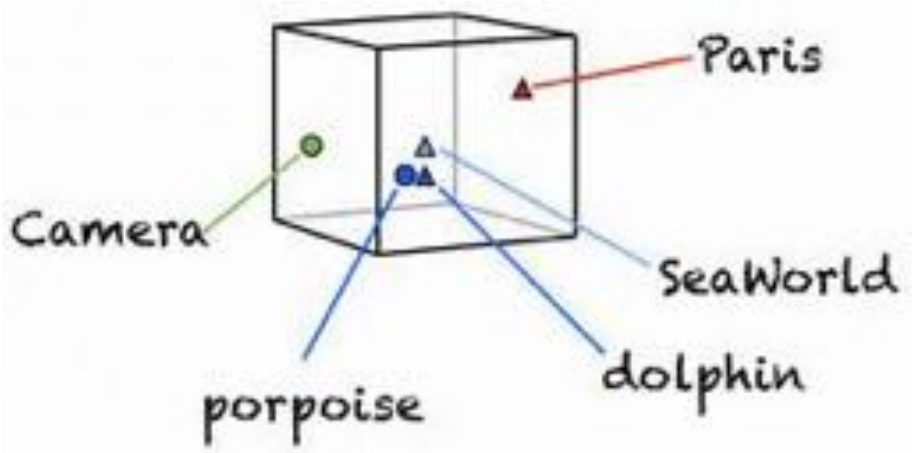
- 乔丹的博士后

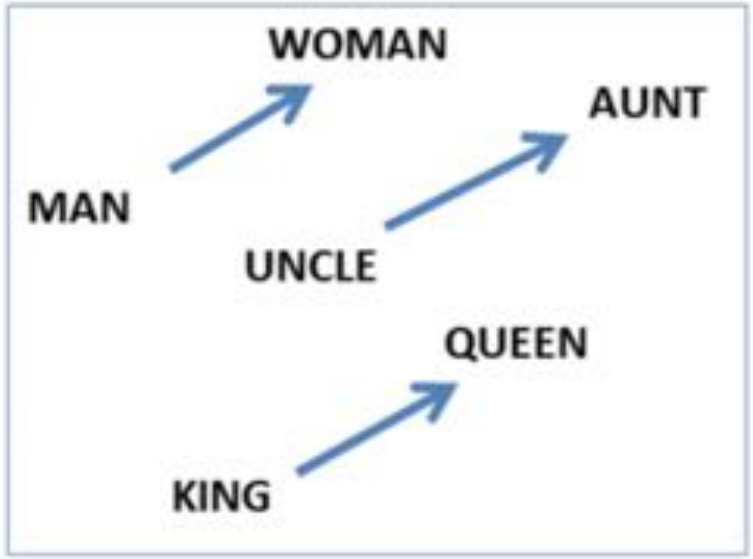
- Word Embedding
- Generative adversarial networks



词嵌入 Word Embedding

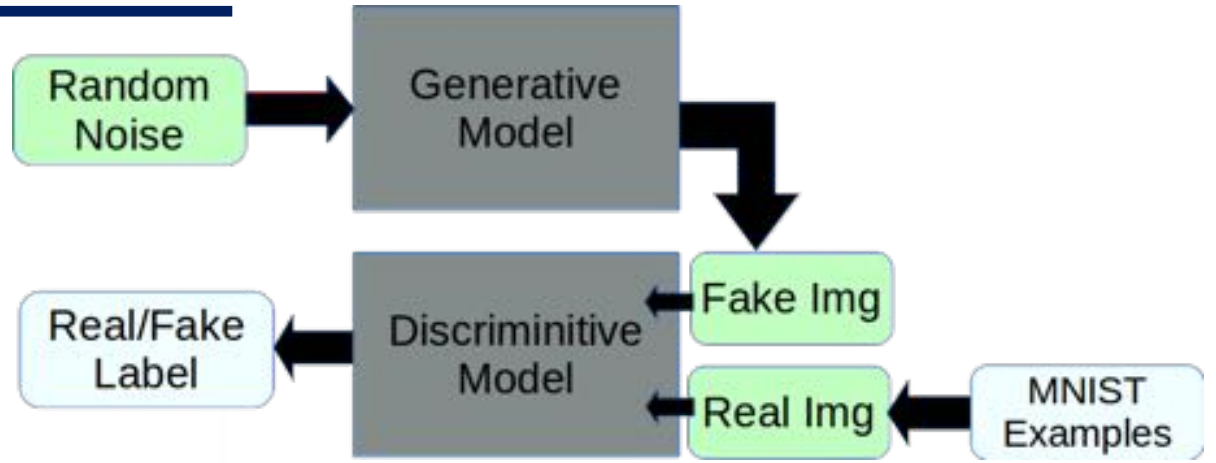


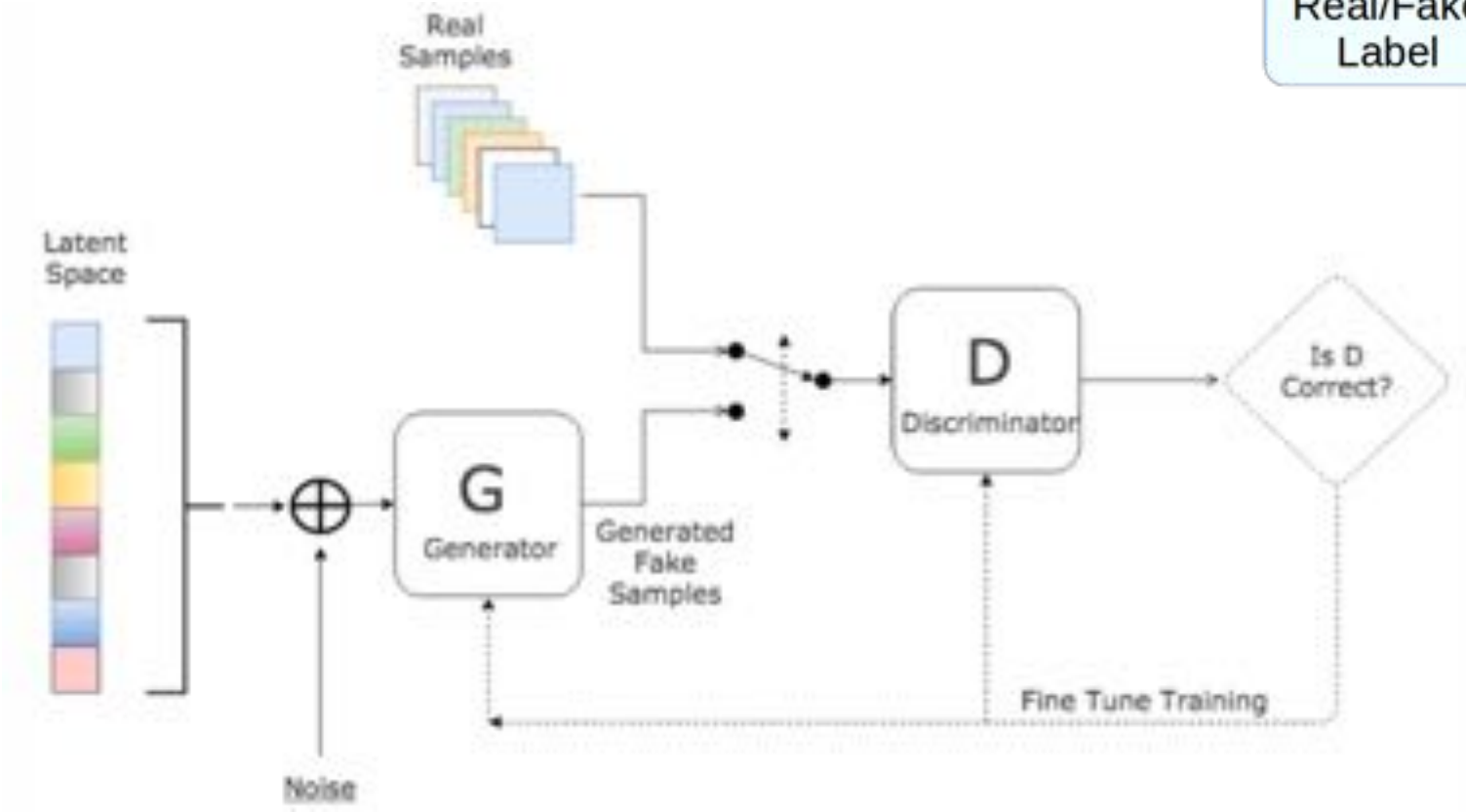




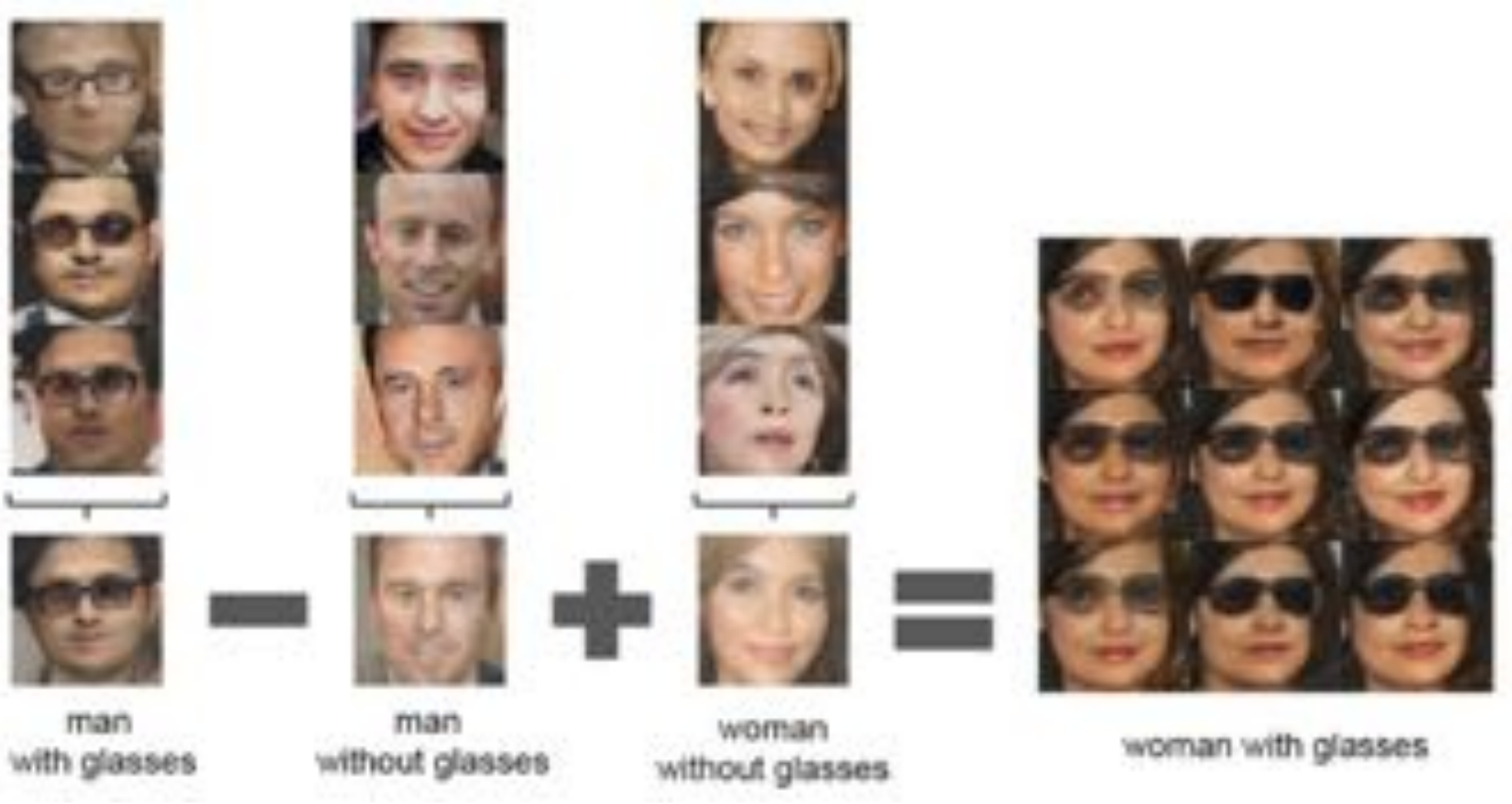
对抗网络 GAN

- Generative adversarial networks





图生成 – GAN应用

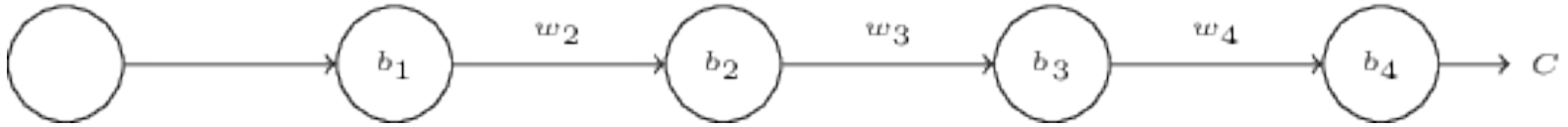

man
with glasses
man
without glasses
woman
without glasses
woman with glasses

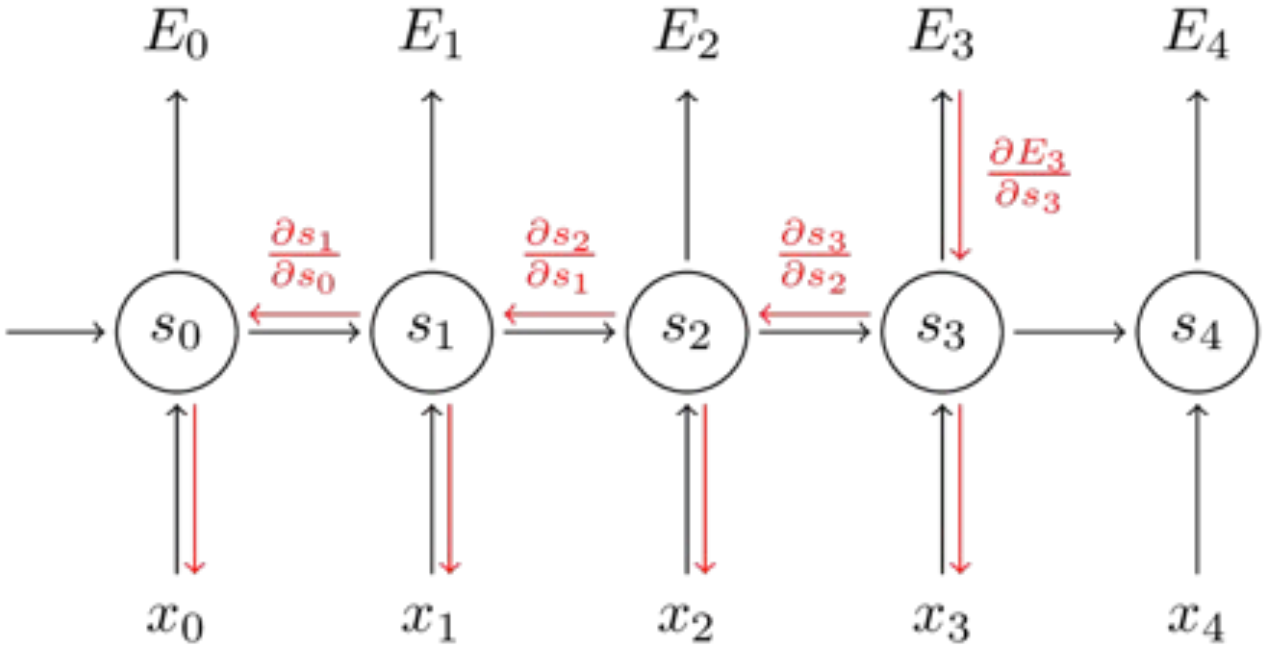
WHY NOW?

- 为什么深度以前不行？

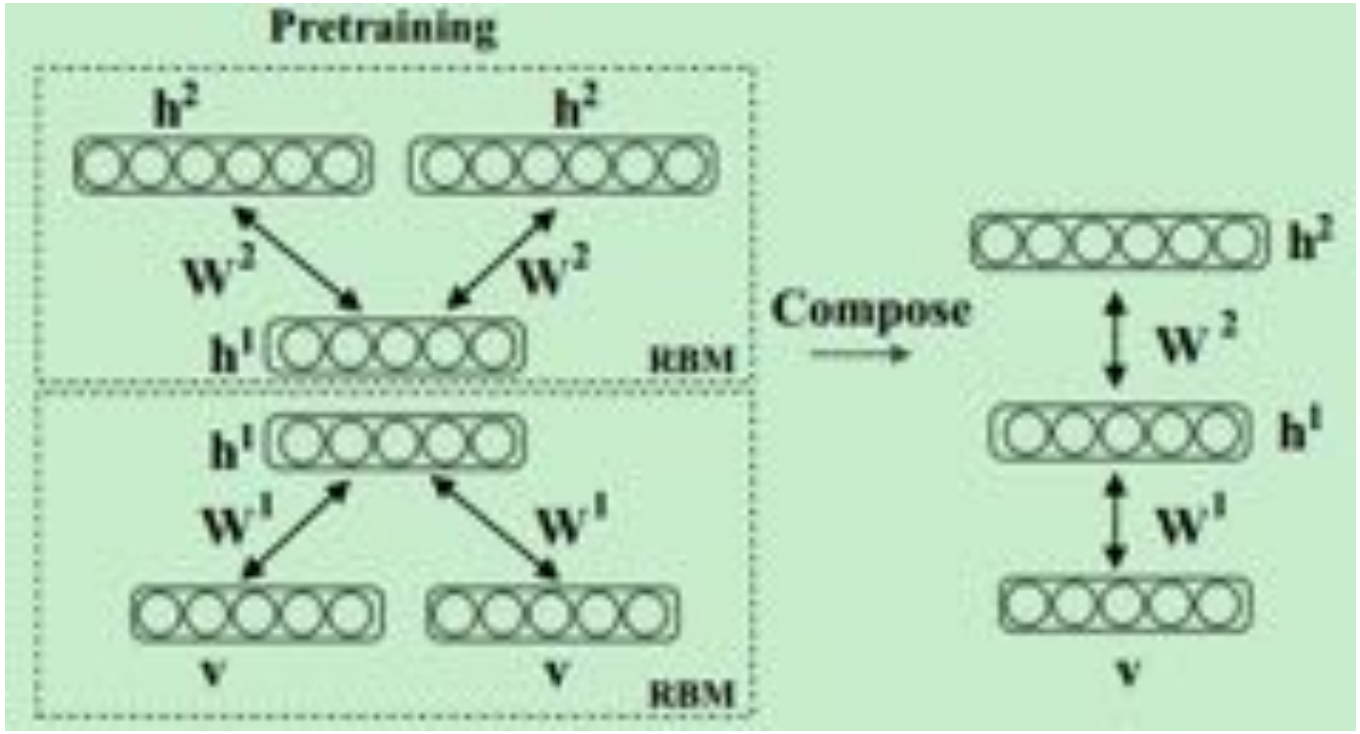
梯度衰退 Gradient Vanishing

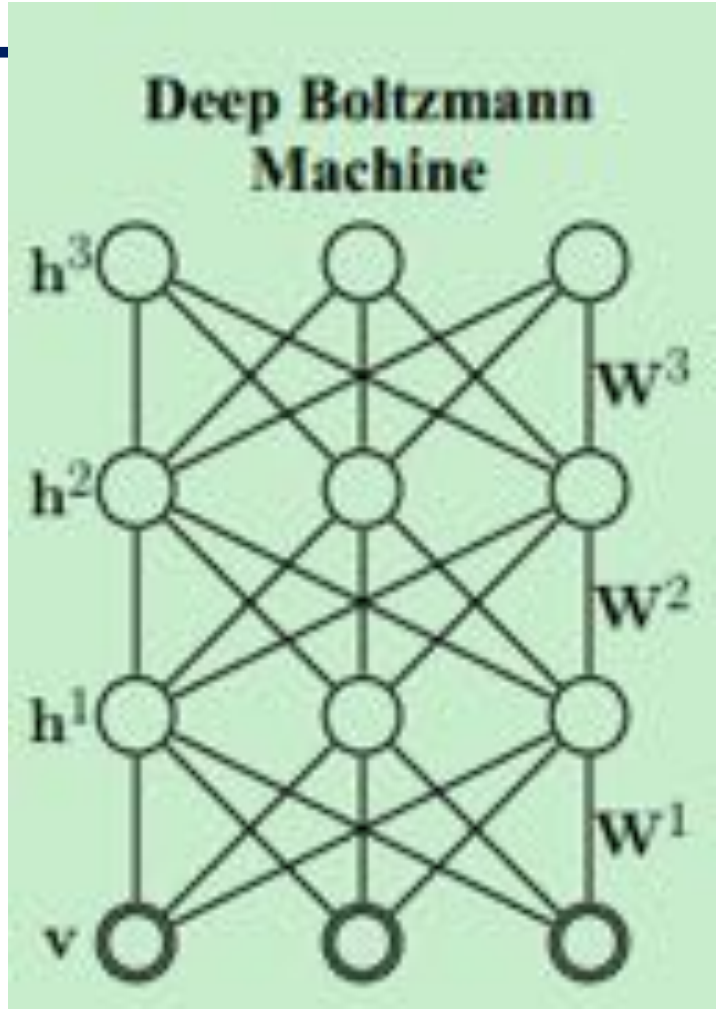
$$\frac{\partial C}{\partial b_1} = \sigma'(z_1) \times w_2 \times \sigma'(z_2) \times w_3 \times \sigma'(z_3) \times w_4 \times \sigma'(z_4) \times \frac{\partial C}{\partial a_4}$$





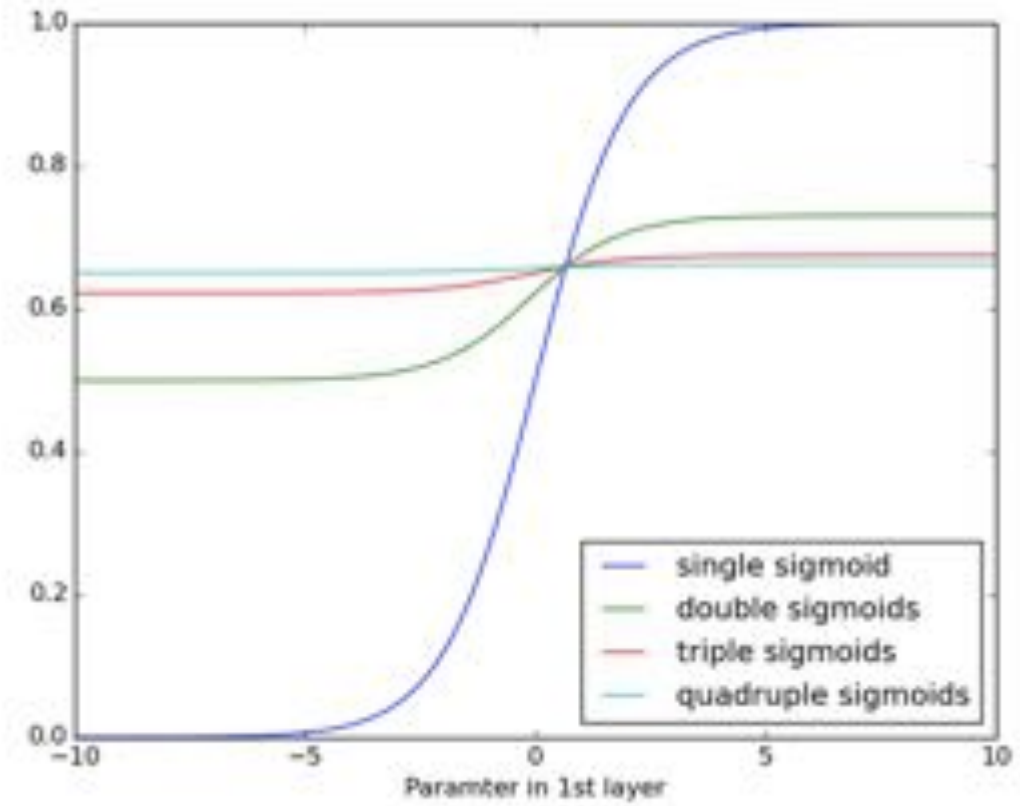
DBM:Pretraining

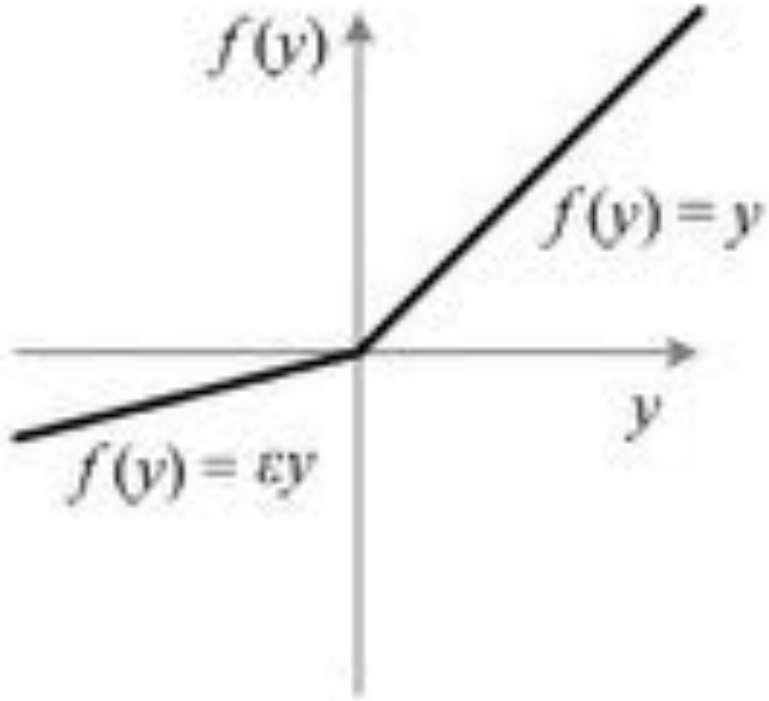




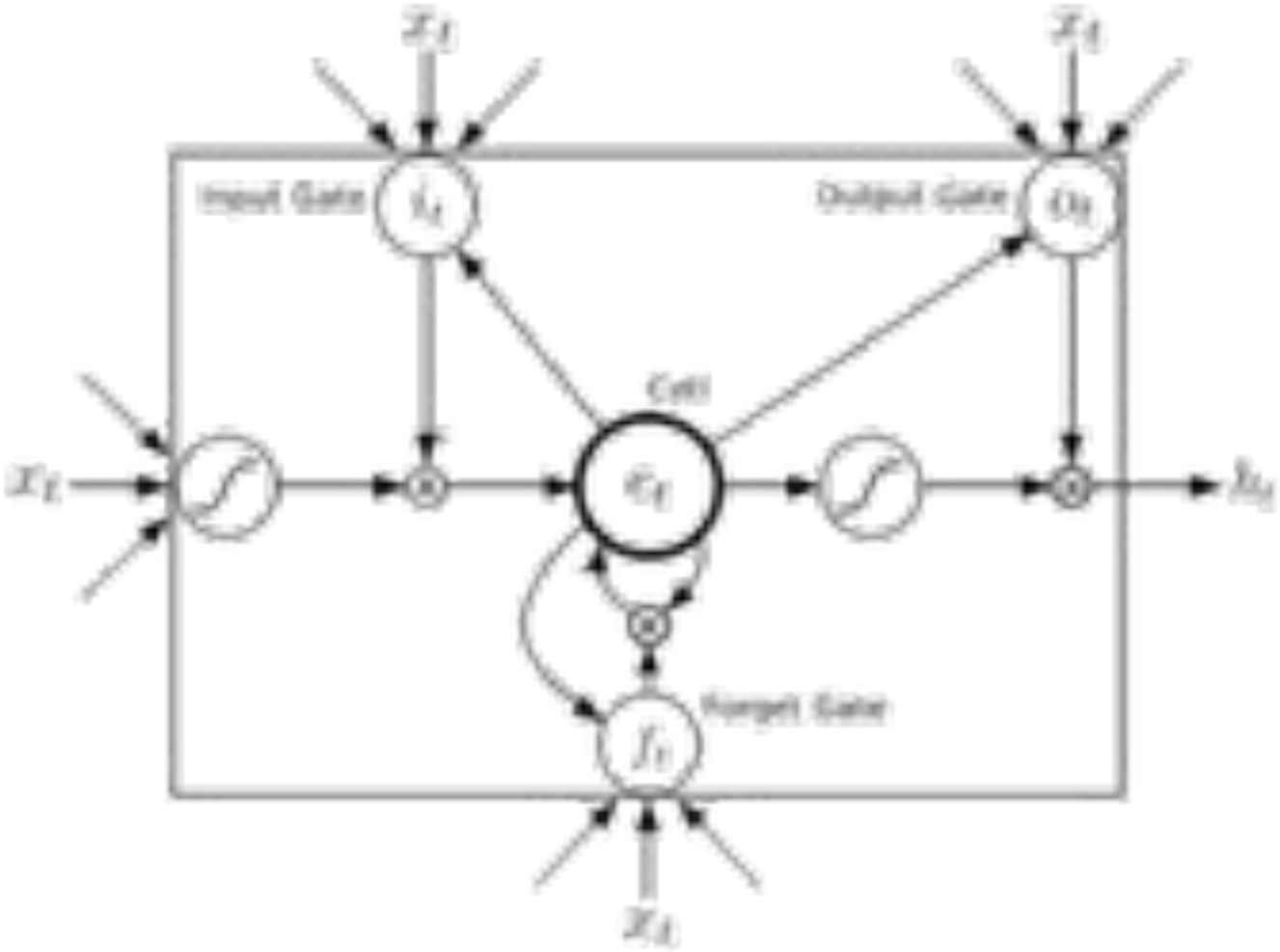
CNN: ReLU

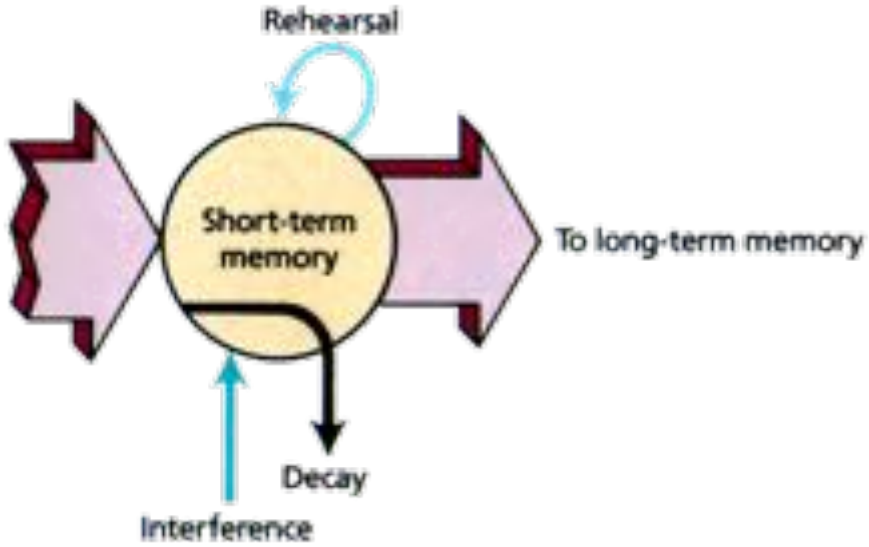
- Rectified Linear Unit





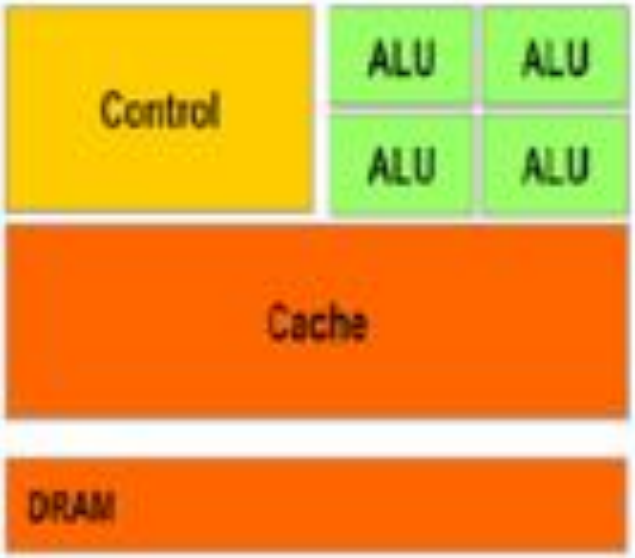
RNN: Forget Gate



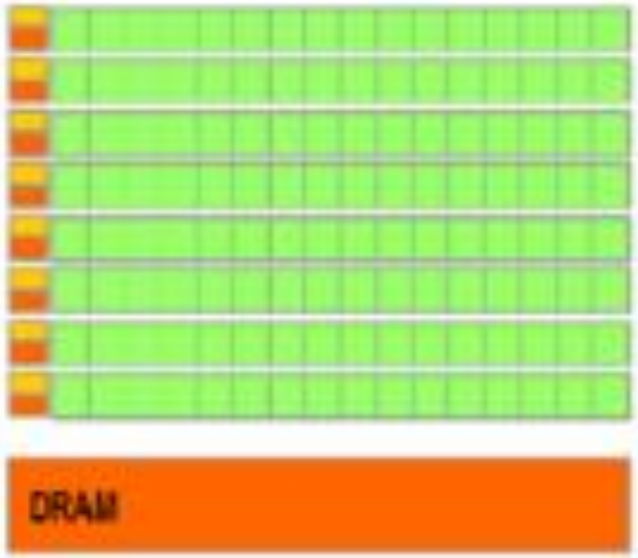


史春奇

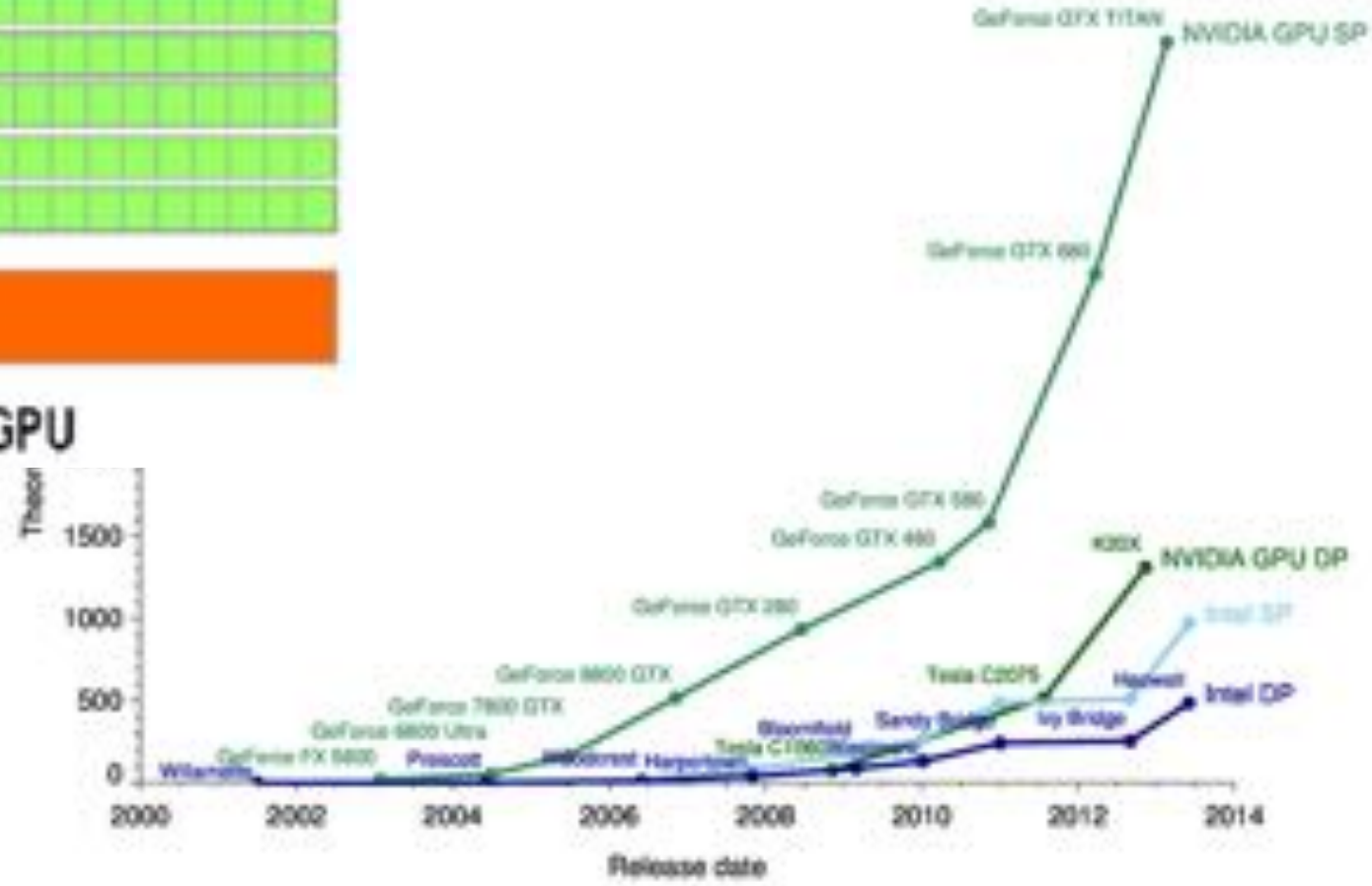
GPU 应用



CPU



GPU

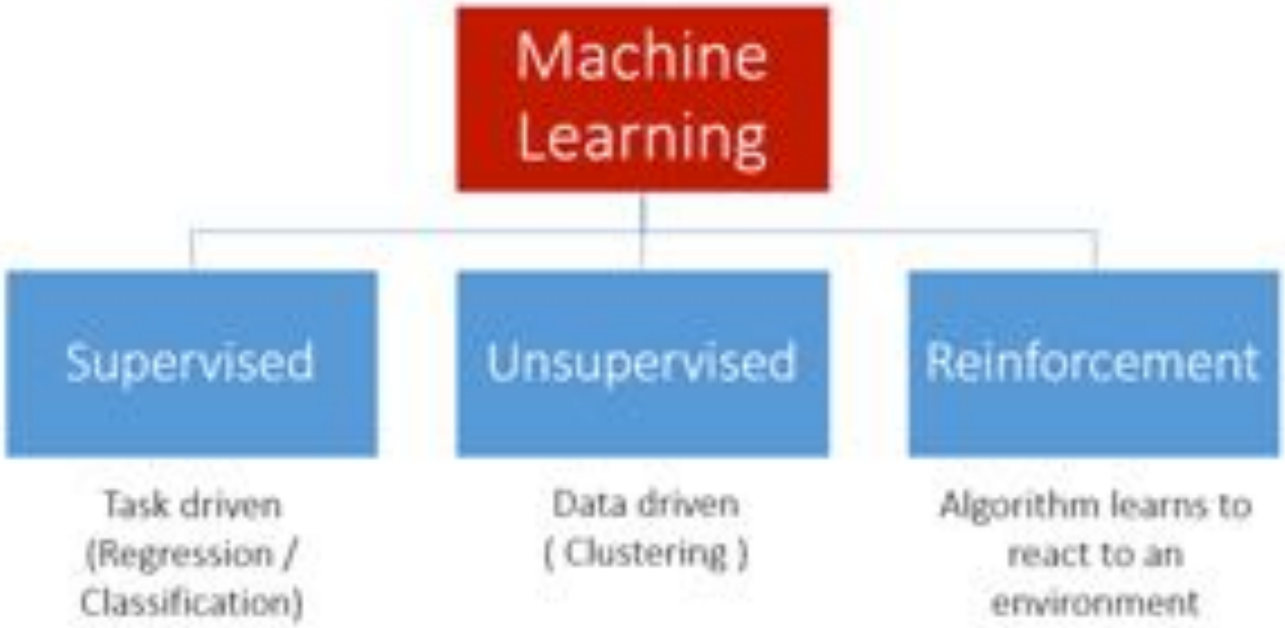


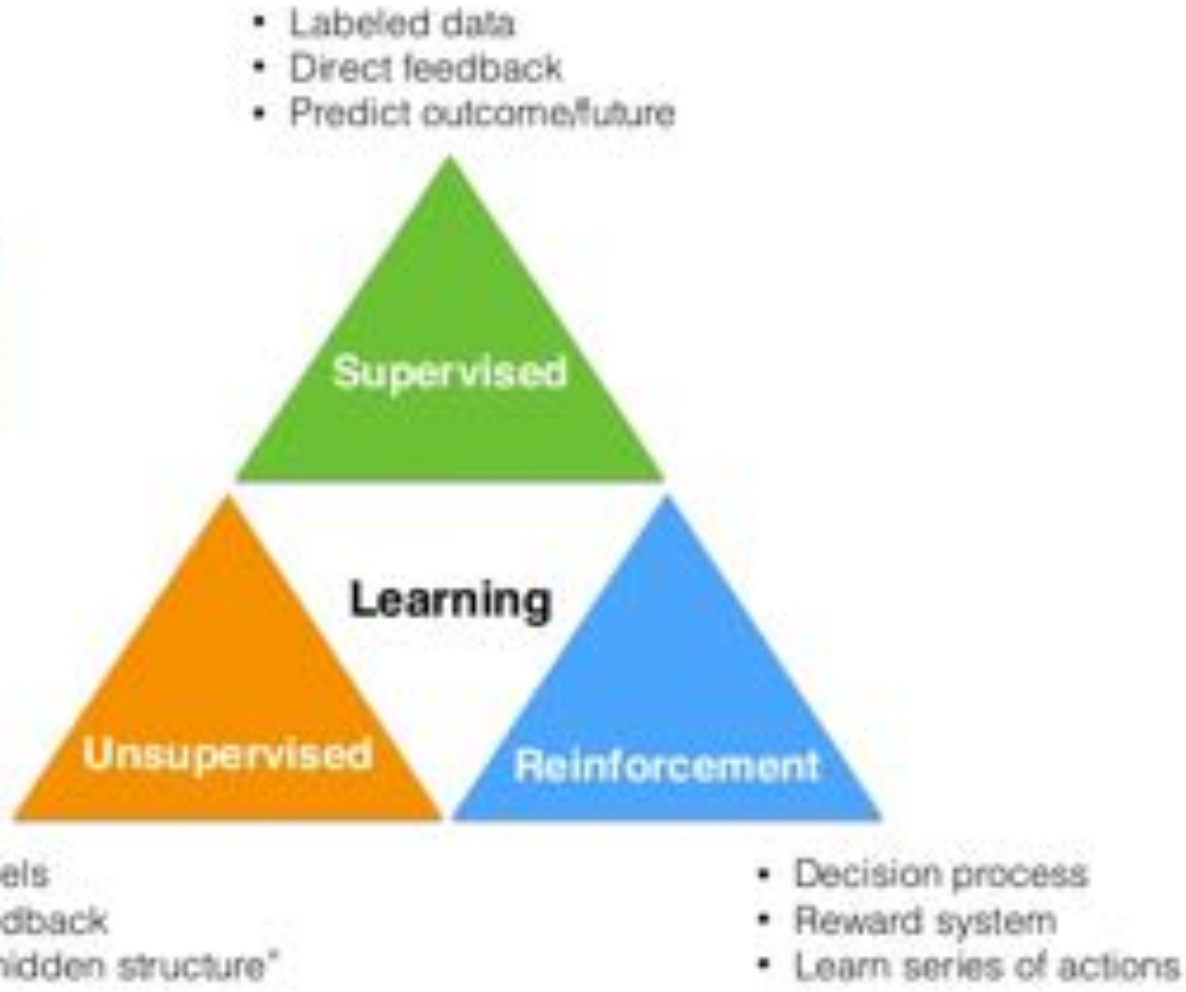
WHAT

- 机器学习包含哪些内容？

史春奇

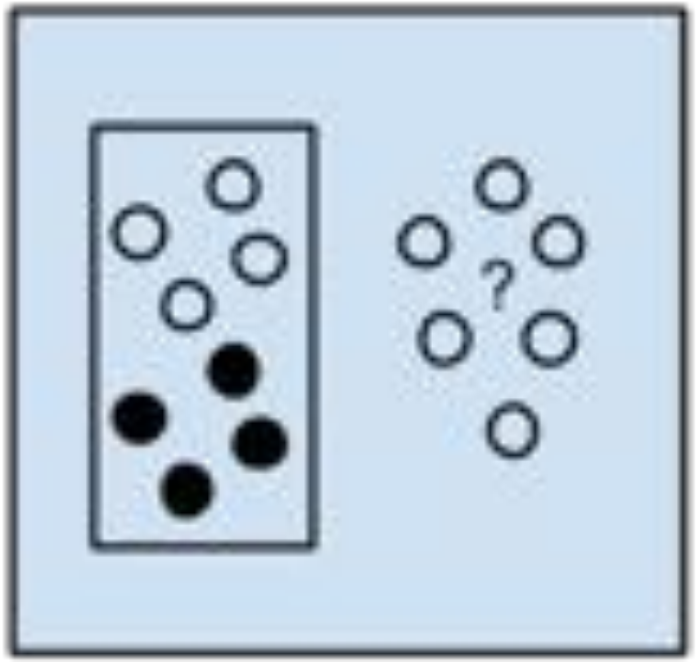
机器学习


Machine Learning
Supervised
Unsupervised
Reinforcement
Task driven (Regression / Classification)
Data driven (Clustering)
Algorithm learns to react to an environment


Labeled data
Direct feedback
Predict outcome/future
Supervised
Learning
Unsupervised
Reinforcement
No labels
No feedback
"Find hidden structure"
Decision process
Reward system
Learn series of actions

监督学习

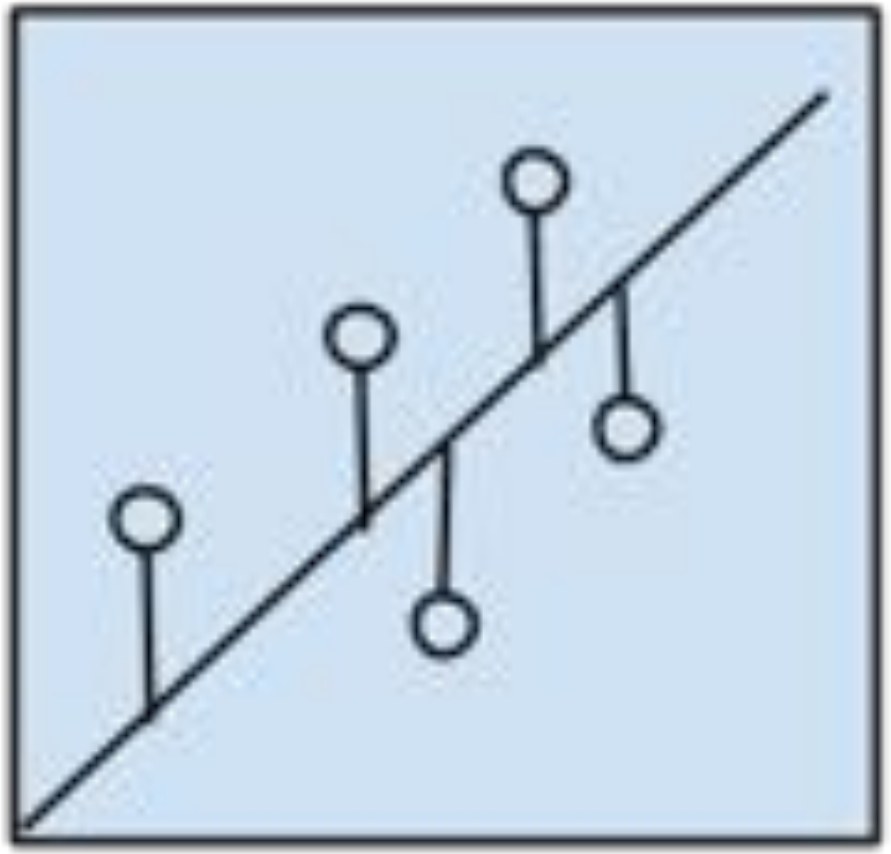
- **Supervised Learning**



Supervised Learning
Algorithms

回归

- **Regression Algorithms**
  - Ordinary Least Squares Regression (OLSR)
  - Linear Regression
  - Logistic Regression
  - Stepwise Regression
  - Multivariate Adaptive Regression Splines (MARS)
  - Locally Estimated Scatterplot Smoothing (LOESS)



Regression Algorithms

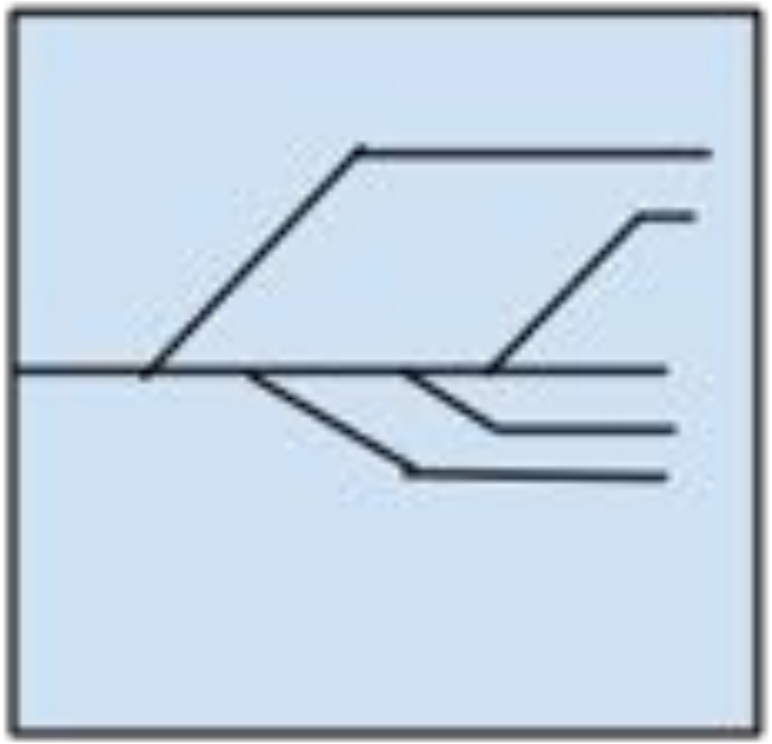
实例算法

- **Instance-based Algorithms**
  - k-Nearest Neighbor (kNN)
  - Learning Vector Quantization (LVQ)
  - Self-Organizing Map (SOM)
  - Locally Weighted Learning (LWL)

Instance-based Algorithms

正则化算法

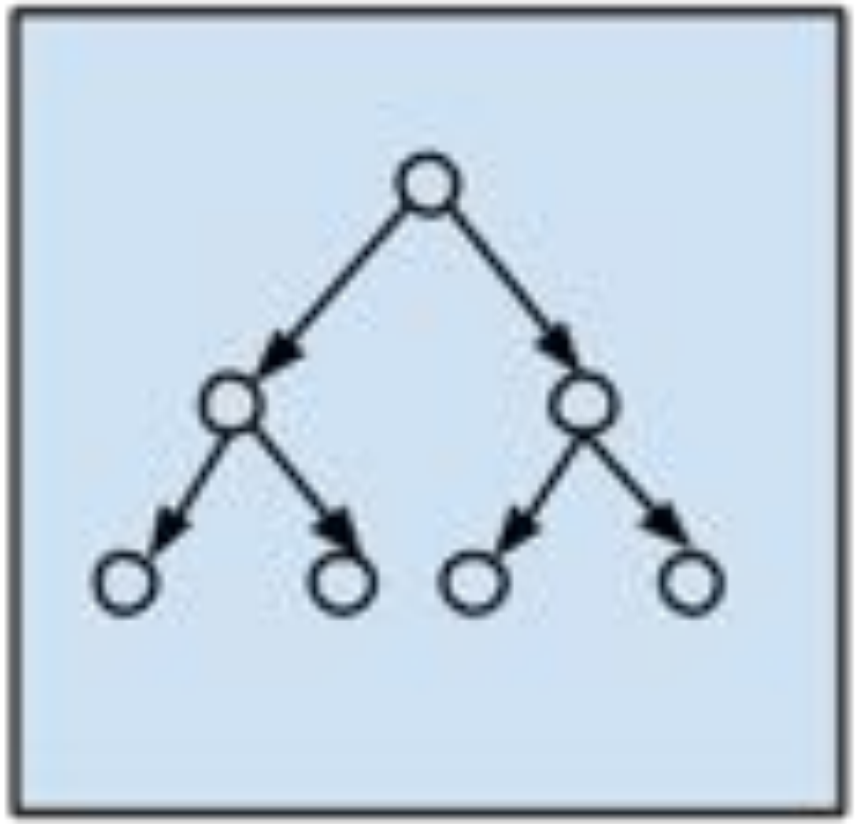
- **Regularization Algorithms**
  - Ridge Regression
  - Least Absolute Shrinkage and Selection Operator (LASSO)
  - Elastic Net
  - Least-Angle Regression (LARS)



Regularization Algorithms

决策树算法

- **Decision Tree Algorithms**
  - Classification and Regression Tree (CART)
  - Iterative Dichotomiser 3 (ID3)
  - C4.5 and C5.0 (different versions of a powerful approach)
  - Chi-squared Automatic Interaction Detection (CHAID)
  - Decision Stump
  - M5
  - Conditional Decision Trees



Decision Tree Algorithms

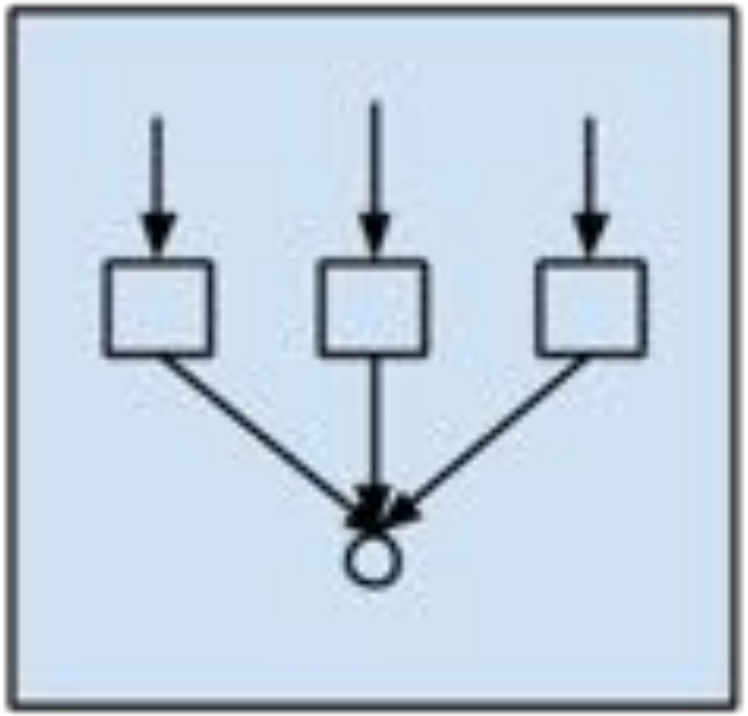
贝叶斯算法

- **Bayesian Algorithms**
  - Naive Bayes
  - Gaussian Naive Bayes
  - Multinomial Naive Bayes
  - Averaged One-Dependence Estimators (AODE)
  - Bayesian Belief Network (BBN)
  - Bayesian Network (BN)

Bayesian Algorithms

集成学习

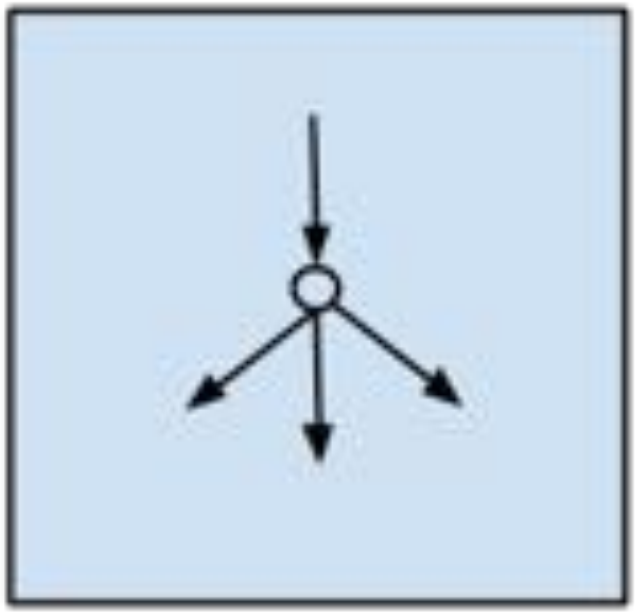
- **Ensemble Algorithms**
  - Boosting
  - Bootstrapped Aggregation (Bagging)
  - AdaBoost
  - Stacked Generalization (blending)
  - Gradient Boosting Machines (GBM)
  - Gradient Boosted Regression Trees (GBRT)
  - Random Forest



Ensemble Algorithms

神经网络

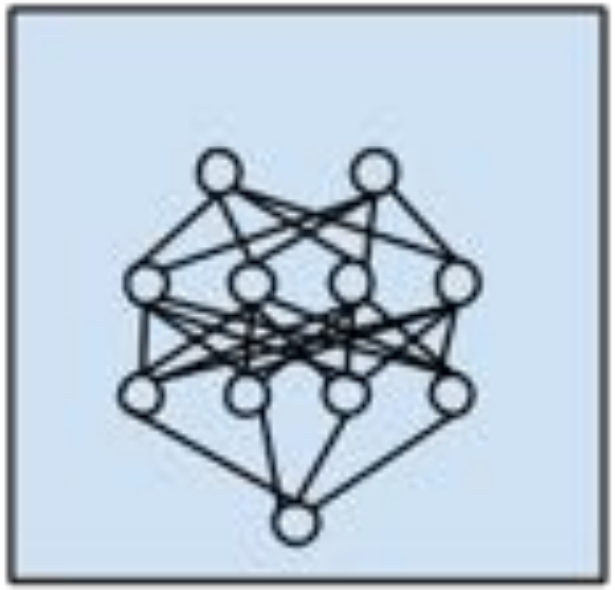
- **Artificial Neural Network Algorithms**
  - Perceptron
  - Back-Propagation
  - Hopfield Network
  - Radial Basis Function Network (RBFN)



Artificial Neural Network Algorithms

深度学习

- **Deep Learning Algorithms**
  - Deep Boltzmann Machine (DBM)
  - Deep Belief Networks (DBN)
  - Convolutional Neural Network (CNN)
  - Stacked Auto-Encoders



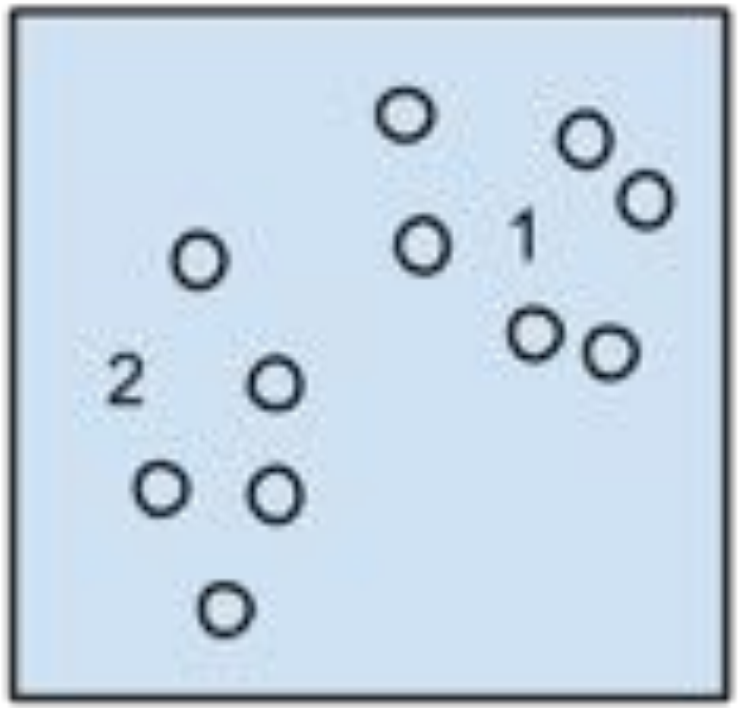
Deep Learning
Algorithms

无监督学习

- **Unsupervised Learning**

聚类算法

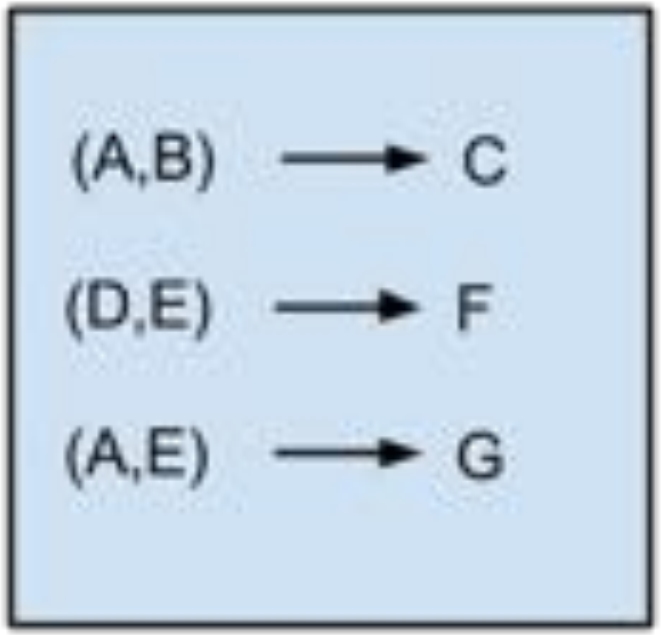
- **Clustering Algorithms**
  - k-Means
  - k-Medians
  - Expectation Maximisation (EM)
  - Hierarchical Clustering



Clustering Algorithms

关联规则

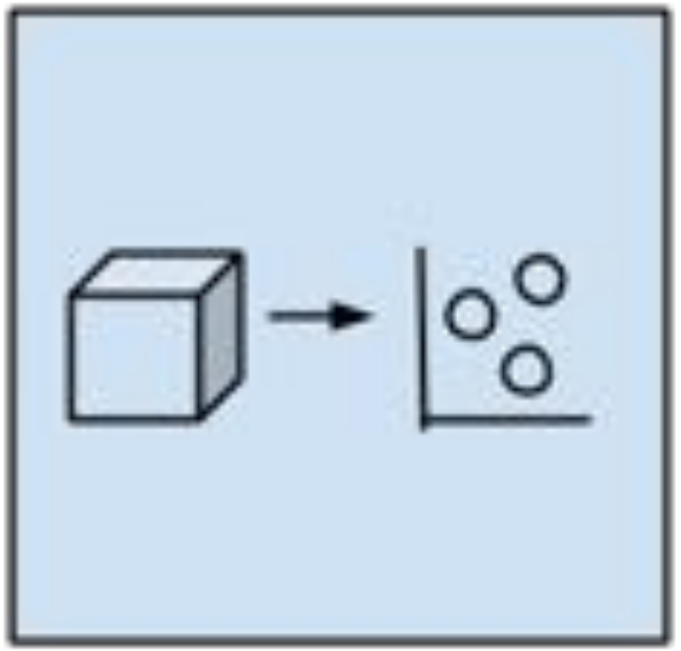
- **Association Rule Learning Algorithms**
  - Apriori algorithm
  - Eclat algorithm



Association Rule Learning Algorithms

降维

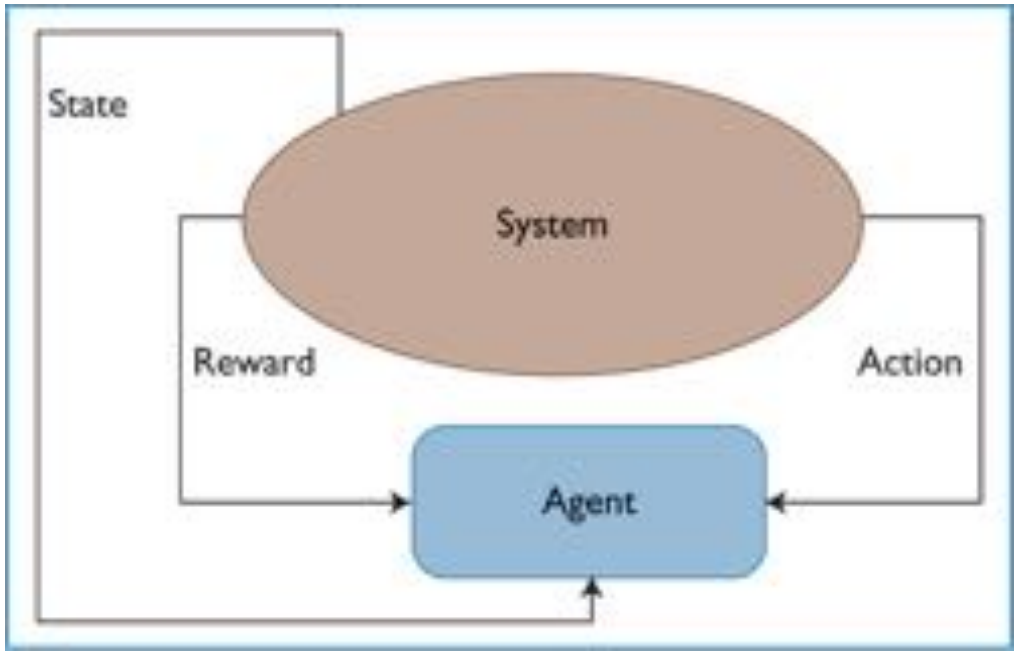
- **Dimensionality Reduction Algorithms**
  - Principal Component Analysis (PCA)
  - Principal Component Regression (PCR)
  - Partial Least Squares Regression (PLSR)
  - Sammon Mapping
  - Multidimensional Scaling (MDS)
  - Projection Pursuit
  - Linear Discriminant Analysis (LDA)
  - Mixture Discriminant Analysis (MDA)
  - Quadratic Discriminant Analysis (QDA)
  - Flexible Discriminant Analysis (FDA)

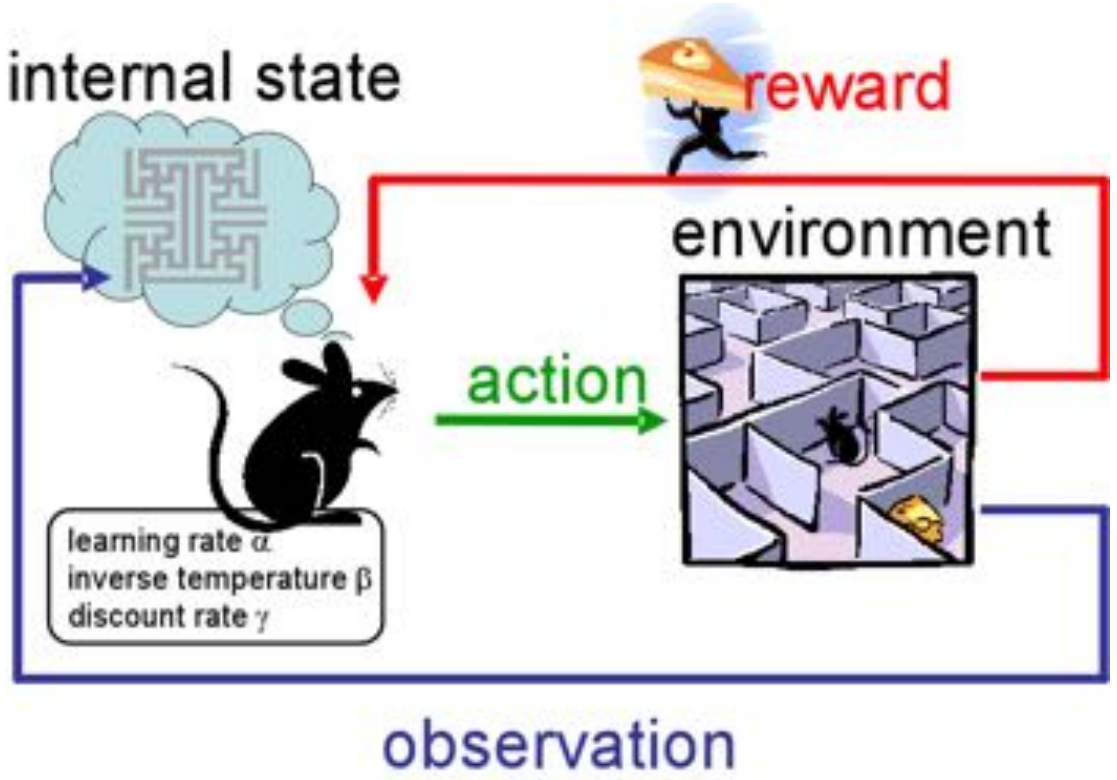


Dimensional Reduction Algorithms

强化学习

- Reinforcement Learning





Reference

- https://badripatro.wordpress.com/2017/01/18/deep-learning-and-machine-learning/
- http://machinelearningmastery.com/a-tour-of-machine-learning-algorithms/
- http://statistics.berkeley.edu/history
- http://zhiyuan.sjtu.edu.cn/teachers/computer

THE END !

- 谢谢您 ~