

# 机器学习

2019年春季

涂仕奎

计算机科学与工程系

[tushikui@sjtu.edu.cn](mailto:tushikui@sjtu.edu.cn)

# 课程介绍

- **48 节课，1-16周，3 学分**

- **1-16周**：星期二 第1节--第2节，东中院3-103
- **1-08周**：星期四 第3节--第4节，东中院3-103

- **目标：**

- 掌握机器学习的基本问题定义、基本模型，对机器学习学科有概览性的认识
- 了解机器学习的前沿技术与研究现状

- **教学方法：**

- 以课堂教学为主，结合阅读自学和大作业，进行分析和讨论

上课节数	上课时间
第一节	8:00 – 8:45
第二节	8:55 – 9:40
第三节	10:00 – 10:45
第四节	10:55 – 11:40

# 课程资料

- Chris Bishop. “Pattern Recognition and Machine Learning” . Springer 2006.
- 李航《统计学习方法》清华大学出版社，2012.
- 周志华《机器学习》清华大学出版社，2016.
- Ian Goodfellow, Yoshua Bengio, Aaron Courville. “Deep Learning” , The MIT Press,2016
- Tom Mitchell. “Machine Learning” . McGraw-Hill, 1997
- Jerome H. Friedman, Robert Tibshirani, and Trevor Hastie. “The Elements of Statistical Learning” . Springer 2004.

# 考核及成绩评定方式

- **平时成绩 : 30%**

- 课堂表现、平时作业，主要考核对知识点的掌握程度。

- **大作业 : 20%**

- 书面报告 + 上机，主要考核独立解决问题的能力
- 深度学习在AI硬件平台的实践

- **期末考试 : 50%**

- 笔试，主要考核对机器学习的基本思想和方法的掌握程度。

# 主讲老师

- 姓名 : 涂仕奎
- Email : [tushikui@sjtu.edu.cn](mailto:tushikui@sjtu.edu.cn)
- 主页 : <http://www.cs.sjtu.edu.cn/~tushikui/>
- 实验室 : 电院3号楼508室 ( CMaCH )
- 办公室 : 电院3号楼541室
- 研究方向 : 机器学习 , 生物信息学



# 助教

- 黄文静 : [huangwenjing@sjtu.edu.cn](mailto:huangwenjing@sjtu.edu.cn)



- 邢智皓 : [xingzhihao@sjtu.edu.cn](mailto:xingzhihao@sjtu.edu.cn)

- 地点 : 电院3号楼508室 ( CMaCH )

- 课程QQ群 : 810 582 120



群名称: CS420-机器学习-2019春

群号: 810582120

- 课程主页 :

<http://cmach.sjtu.edu.cn/course/cs420>

# 校历

1-16周：星期二 第1节--第2节，东中院3-103

1-08周：星期四 第3节--第4节，东中院3-103

## 上海交通大学2018-2019学年校历

	九月		十月		十一月		十二月		一月		二月		三月		四月		五月		六月		七月		八月		九月																							
星期一	10	17	24	1	8	15	22	29	5	12	19	26	3	10	17	24	31	7	14	21	28	4	11	18	25	1	8	15	22	29	6	13	20	27	3	10	17	24	1	8	15	22	29	5	12	19	26	2
星期二	11	18	25	2	9	16	23	30	6	13	20	27	4	11	18	25	1	8	15	22	29	5	12	19	26	2	9	16	23	30	6	13	20	27	3													
星期三	12	19	26	3	10	17	24	31	7	14	21	28	5	12	19	26	2	9	16	23	30	6	13	20	27	3	10	17	24	31	7	14	21	28	4													
星期四	13	20	27	4	11	18	25	1	8	15	22	29	6	13	20	27	3	10	17	24	31	7	14	21	28	4	11	18	25	1	8	15	22	29	5													
星期五	14	21	28	5	12	19	26	2	9	16	23	30	7	14	21	28	4	11	18	25	2	9	16	23	30	6	13	20	27	5	12	19	26	2	9	16	23	30	6									
星期六	15	22	29	6	13	20	27	3	10	17	24	1	8	15	22	29	5	12	19	26	2	9	16	23	30	6	13	20	27	3	#	17	24	31	7													
星期日	16	23	30	7	14	21	28	4	11	18	25	2	9	16	23	30	6	13	20	27	3	#	17	24	3	10	17	24	31	7	14	21	28	4	#	18	25	1	8									
周数	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	1	2	3	4	5	6	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	1	2	3	4	5	6
学期	秋季学期								寒假				春季学期								夏季学期				暑假																							

节假日：（节假日安排按国务院通知执行）

备注：

- |            |                        |              |               |
|------------|------------------------|--------------|---------------|
| 1、 中 秋     | 9月24日（周一）休假            | 7、 清 明       | 4月5日（周五）休假    |
| 2、 国 庆     | 10月1、2、3日（周一、二、三）休假    | 8、 校 庆       | 4月8日（周一）      |
| 3、 元 旦     | 1月1日（周二）休假             | 9、 五 一 劳 动 节 | 5月1日（周三）休假    |
| 4、 春 节     | 2月4、5、6日（除夕、初一、初二）     | 10、 端 午      | 6月7日（周五）休假    |
| 5、 寒 假     | 1月14日至2月24日 6周（其中包括春节） | 11、 暑 假      | 7月29日至9月8日 6周 |
| 6、 研究生毕业典礼 | 3月30日上午                | 12、 本科生毕业典礼  | 7月7日上午        |

说明：1. 秋季学期、春季学期各18周，其中最后2周为考试周；夏季学期为4周；寒假6周；暑假6周。

2. 秋季学期9月8日为新生报到日，9月9日为老生报到日，9月10日正式行课；春季学期2月24日为学生注册报到日，2月25日正式行课。

# **Any Questions?**

# 什么是机器学习？

## Machine Learning



what society thinks I  
do



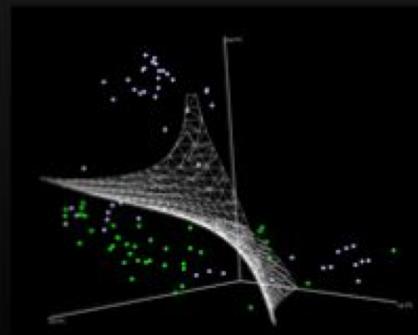
what my friends think  
I do



what my parents think  
I do

$$L_p = \frac{1}{2} \|\mathbf{w}\|^2 - \sum_{i=1}^n \alpha_i y_i (\mathbf{x}_i \cdot \mathbf{w} + b) + \sum_{i=1}^n \alpha_i$$
$$\alpha_i \geq 0, \forall i$$
$$\mathbf{w} = \sum_{i=1}^n \alpha_i \mathbf{x}_i, \sum_{i=1}^n \alpha_i y_i = 0$$
$$\nabla g(\theta_t) = \frac{1}{n} \sum_{i=1}^n \nabla \ell(x_i, y_i; \theta_t) + \nabla r(\theta_t).$$
$$\theta_{t+1} = \theta_t - \eta_p \nabla \ell(x_{i(t)}, y_{i(t)}; \theta_t) - \eta_r \cdot \nabla r(\theta_t)$$
$$\mathbb{E}_{i(t)}[\ell(x_{i(t)}, y_{i(t)}; \theta_t)] = \frac{1}{n} \sum_i \ell(x_i, y_i; \theta_t).$$

what other programmers  
think I do



what I think I do

```
>>> from sklearn import svm
```

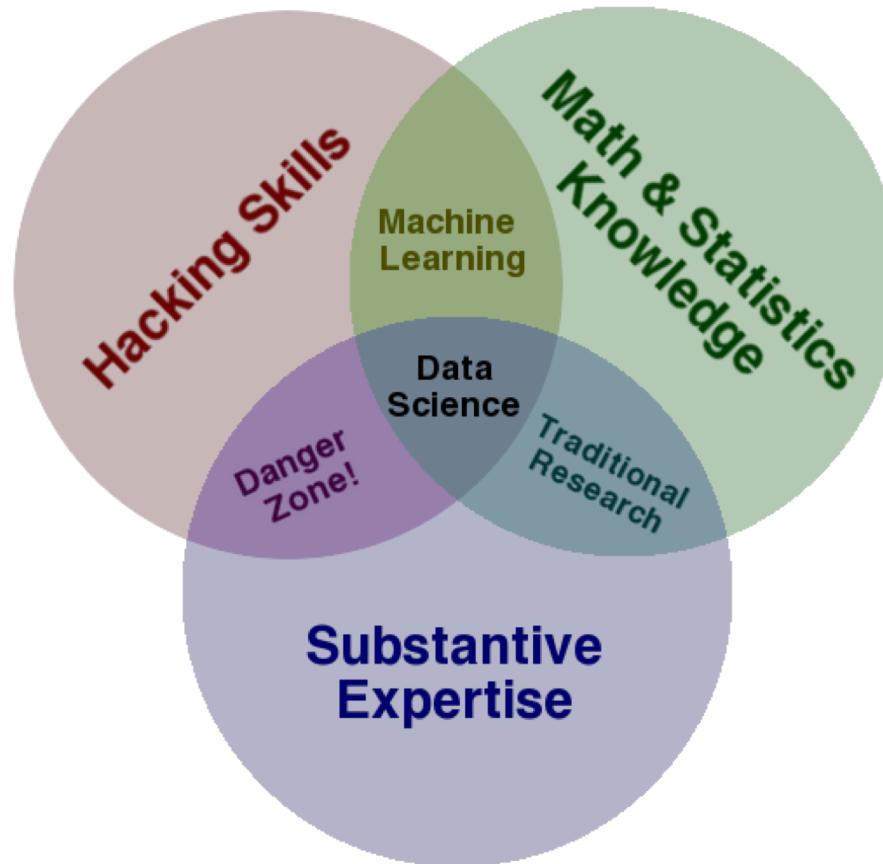
what I really do

# 机器学习的定义

- **Arthur Samuel 的定义(1959)**
  - 在不直接针对问题进行编程的情况下，赋予计算机学习能力的一个研究领域。
- **Tom Mitchell 的定义(1997)**
  - “机器学习这门学科所关注的问题是：计算机程序如何随着经验积累自动提高性能。” (Machine Learning is the study of computer algorithms that improve automatically through experience.)
  - “对于某类任务T和性能度量P，如果一个计算机程序在T上以P衡量的性能随着经验E而自我完善，那么我们称这个计算机程序在从经验E学习。”

# THE DATA SCIENCE VENN DIAGRAM

By Drew Conway

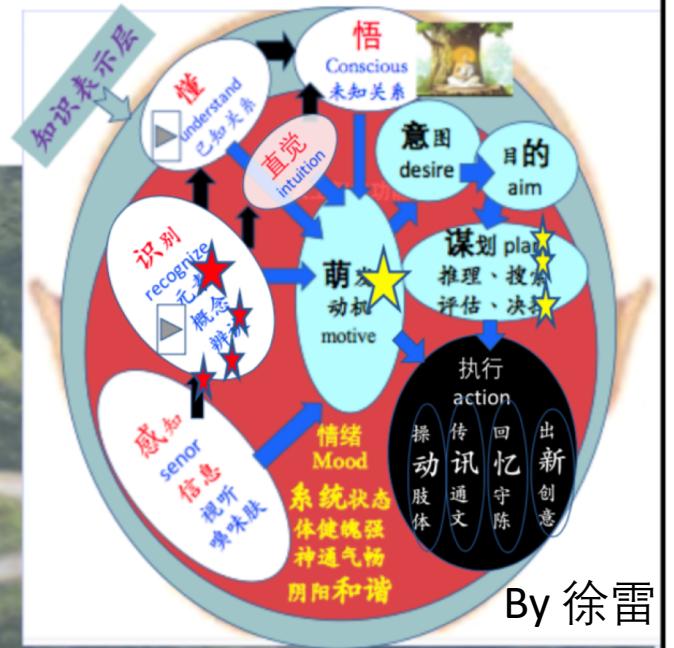


Machine learning = **Hacking skills** + **Math + Statistics**

# 机器学习是人工智能的核心

- Langley ( 1996) 定义的机器学习
  - “机器学习是一门人工智能的科学，该领域的主要研究对象是人工智能，特别是如何在经验学习中改善具体算法的性能” 。(Machine learning is a science of the artificial. The field's main objects of study are artifacts, specifically algorithms that improve their performance with experience.)

# 人工智能第三次浪潮



1950年图灵测试提出后

# 当前背景

- 世界三大尖端科技
  - 20世纪：空间技术、能源技术、**人工智能**
  - 21世纪：**基因工程**、纳米科技、**人工智能**
- 2016年全球投资人工智能：200-300亿美元
- 中国人工智能融资规模约为26亿，全球第二
- 美国、中国、英国位列全球目前人工智能项目排名的前三甲，分别为2905、709、366家。

# Watson 沃森：IBM制造的电脑问答（Q&A）系统

## 认知

听懂语音  
抓住关键词  
理解词间关系

## 思考

搜寻记忆  
证据推理  
选择决断

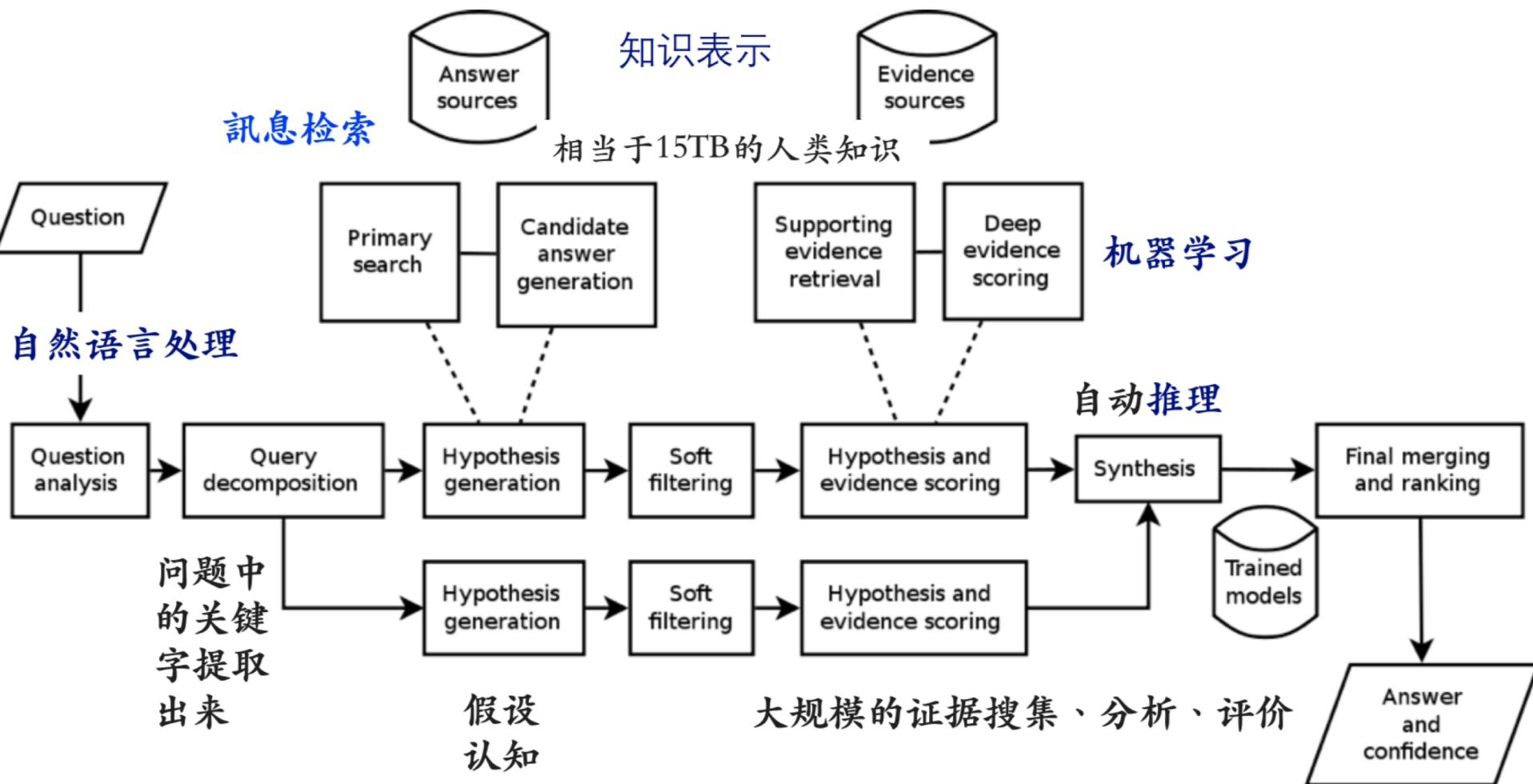
## 通讯

搜肠刮肚  
组织语句  
吐词清楚



2011年，沃森参加综艺节目《危险边缘》来测试它的能力，这是该节目有史以来第一次人与机器对决。打败了最高奖金得主布拉德·鲁特尔和连胜纪录保持者肯·詹宁斯。

彻底搜索这些关键字，它会针对关键字仔细搜索能找到的信息及其上下文、分类名、答案的类型、时间、地点、性别以及其他资料。



The high-level architecture of IBM's DeepQA

# AlphaGo 人机大战

认知  
看懂棋局  
抓住关键

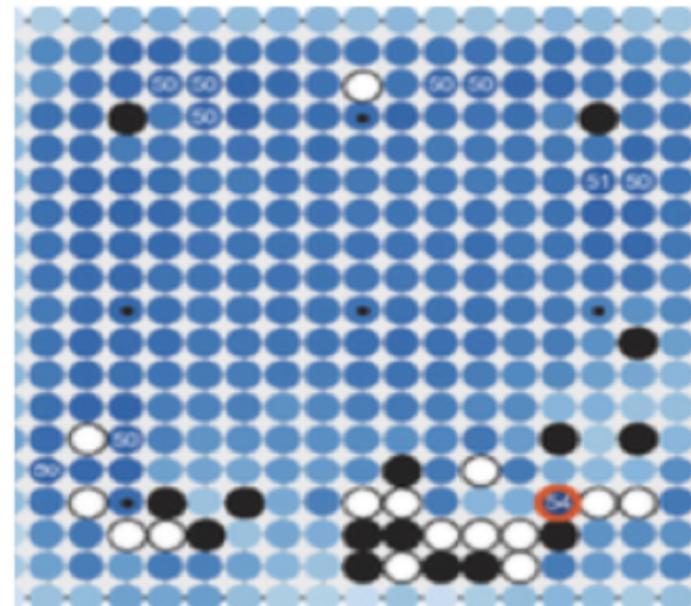
围棋无法仅通过寻找最佳棋步来解决  
游戏一盘平均约有150步  
每一步平均约有200种可选的下法  
需要解决的可能性  $200^{150}$  种

思考

搜寻记忆  
推理赢面  
选择决断

行动

落子到位



# ImageNet Challenge

## 深度学习 驱动识别（标签）全面突破

- 1,000 object classes (categories).
- Images:
  - 1.2 M train
  - 100k test.



监督学习

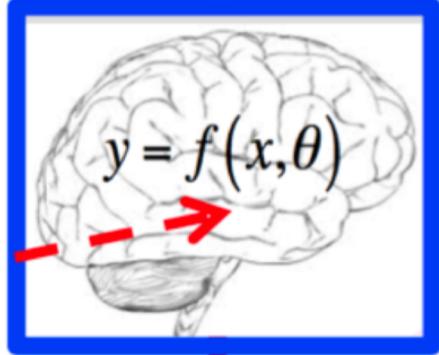


教师

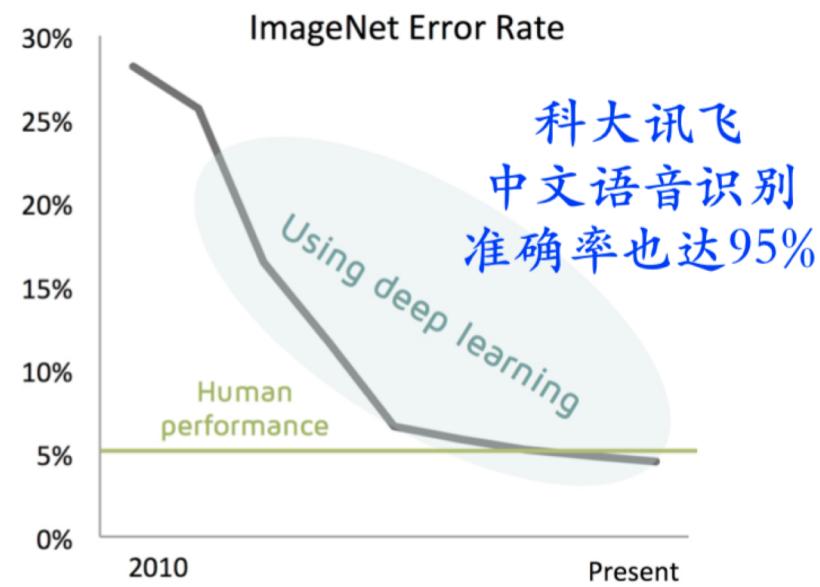
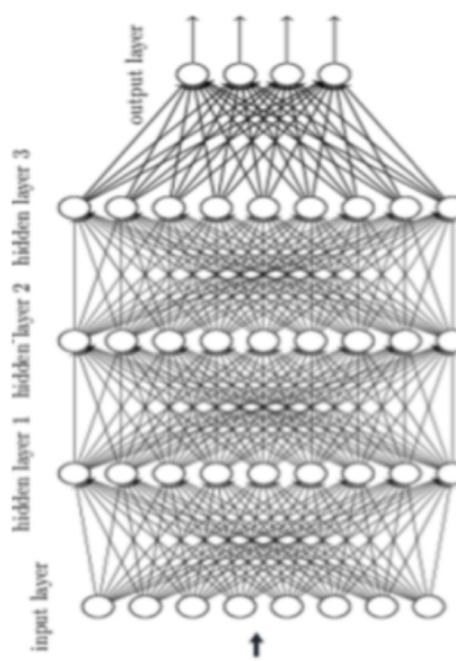
$$e = t - y \text{ 误差}$$



输出



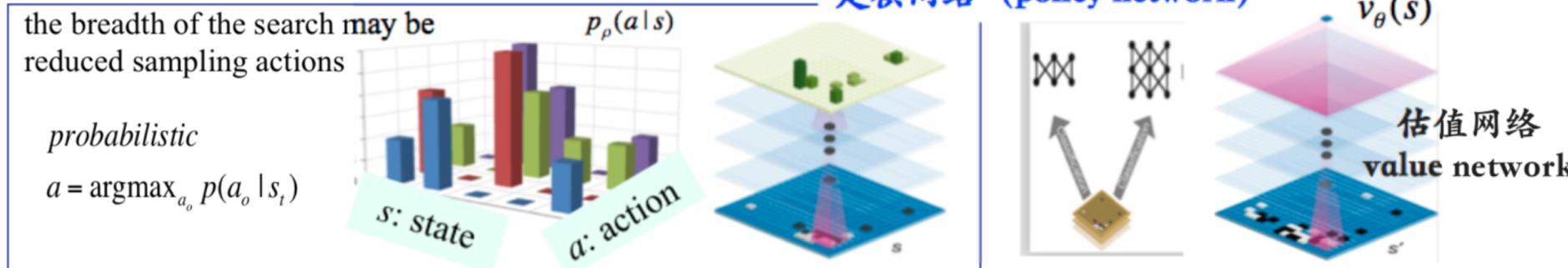
输入



科大讯飞  
中文语音识别  
准确率也达95%

达到人类水平，进入实用阶段！

- 它利用大数据（3000万步棋着）通过深度网络学习得出对当前棋局好坏的经验评估和下一步各种落子的**经验落子概率分布**；



- 还利用大数据通过浅层网络学习得出快速走子策略，驱动蒙特卡洛树搜索侦察，获得获胜机会的前瞻评估 $z_L$ 。

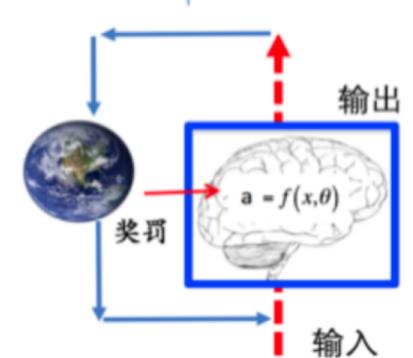
综合经验评估和前瞻评估  $V(s_L) = (1 - \lambda)v_\theta(s_L) + \lambda z_L$

- 代表行为主义的Q学习，进一步考虑**得分奖励**得到**改进评估**。

$$Q(s_t, a) = Q^{old}(s_t, a)(1 - \alpha_t) + \alpha_t[R_{t+1} + \gamma V^{old}(s_{t+1})]$$

$$Q(s_t, a) = Q^{old}(s_t, a)(1 - \alpha_t) + \alpha_t V(s_L)$$

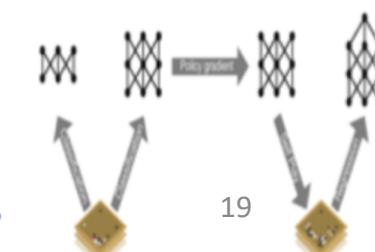
增强学习



- 改进后评估再综合经折算后**经验落子概率**才真正落子

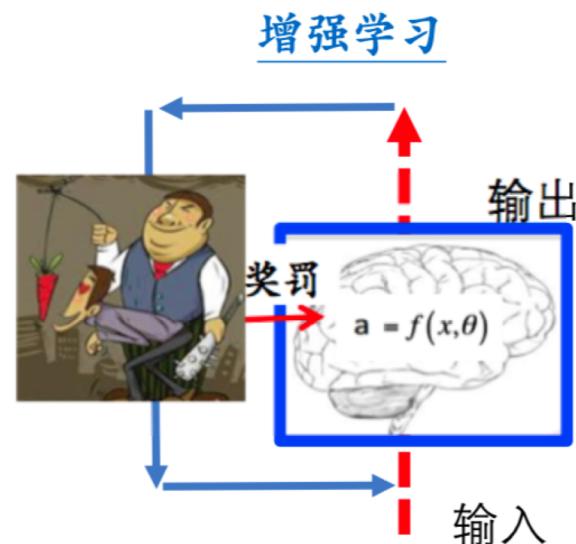
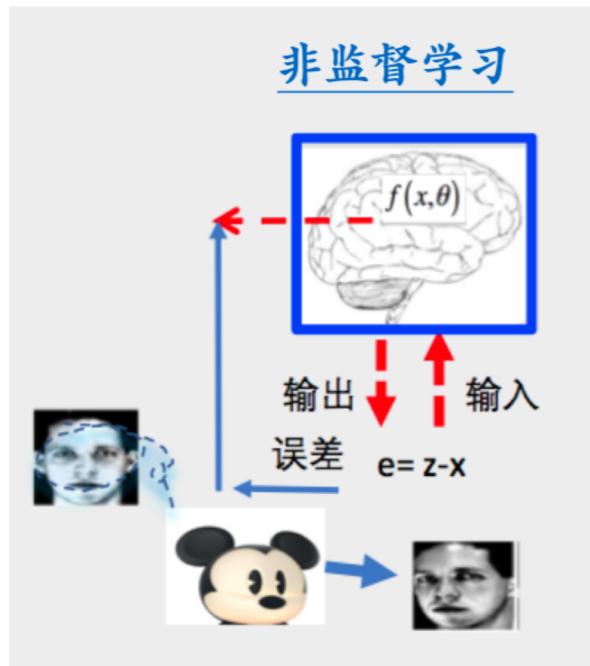
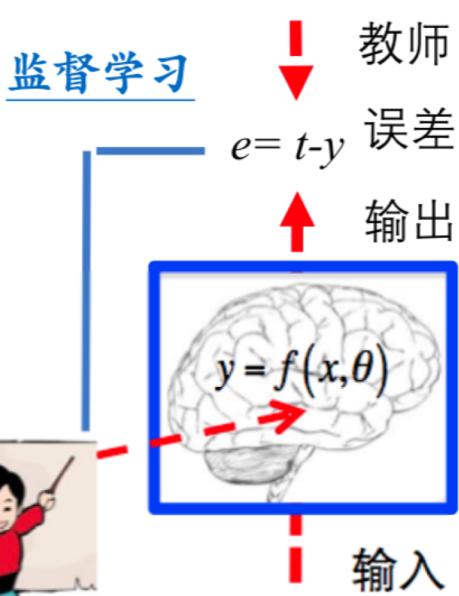
$$a = \operatorname{argmax}_{a_o} [Q(s, a_o) + \frac{p(a_o | s_t)}{1 + N(s, a)}]$$

除以该走子的被用频率，鼓励奇招。



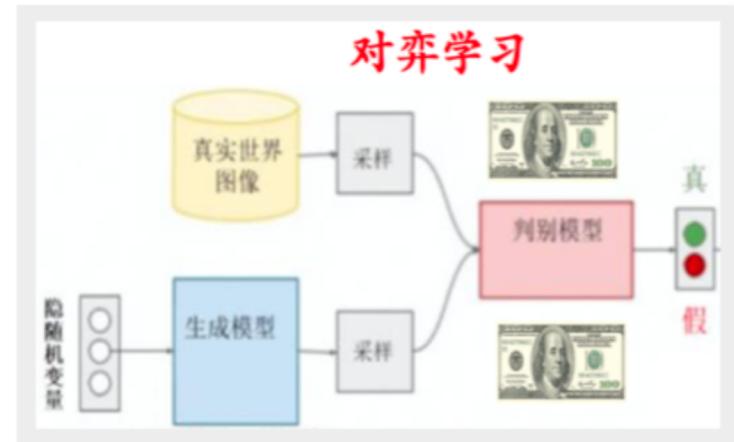
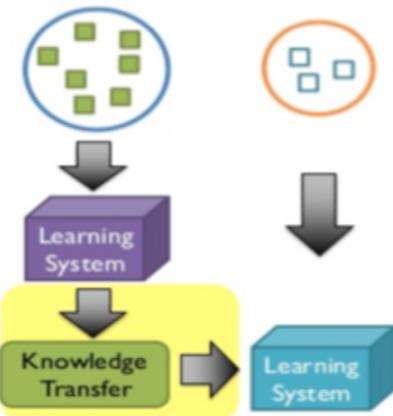
# 七个基本学习机制

Slide from Lei Xu



## 半监督、迁移学习

Transfer Learning



## 群体智能 (crowd intelligence)



Integrated learning 集成学习



# 中国人工智能第一次浪潮：70年代末至80年代末

先五位百岁级先驱、中科院学部委员



钱学森先生  
(1911.12生)

提倡思维科学，  
尤其是形象思维。

王湘浩先生  
(1915.5生)

80年在吉大办了最早  
全国性的人工智能讨  
论班。注重符号系统，  
研究定理机器证明

吴文俊先生  
(1919.5生)

数学机械化方法、  
几何定理自动证  
明、不同于符号  
方法的新方法

程民德先生  
(1917.1生)

**常先生领衔并与程先生搭档，  
一起领导推动认知智能（模  
式识别、信号与图像处理、  
机器视觉）的发展，以及第  
一批三个智能科学的国家重  
点实验室建立。**

# 中国人工智能产业生态图谱

iiMedia 艾媒咨询  
Research

基础层



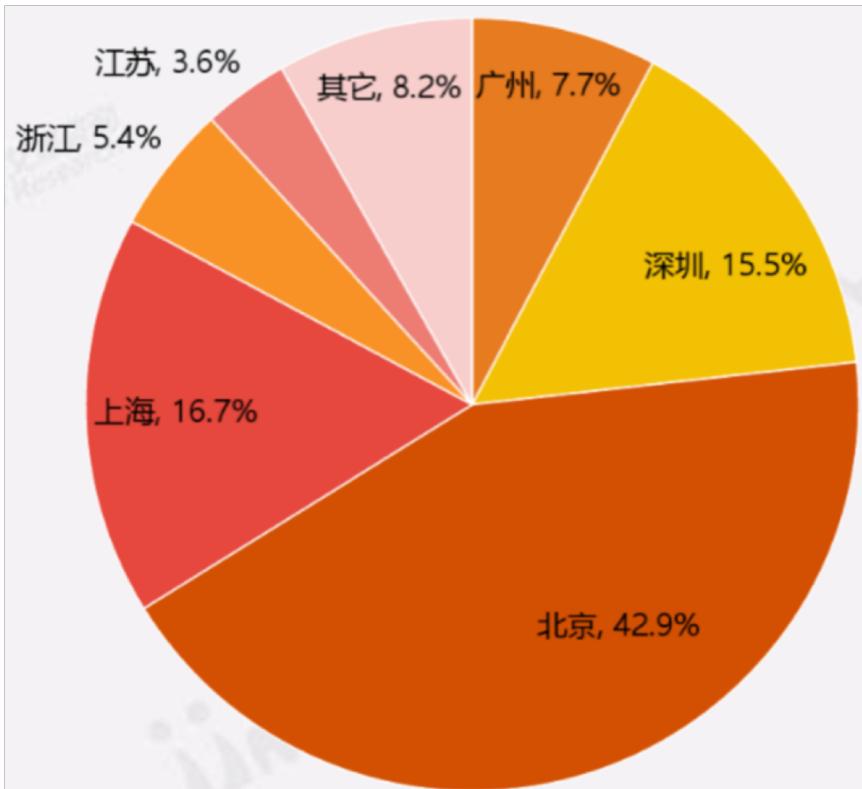
技术层



应用层



## 中国人工智能创业公司所在地区分布



注：截至2017年1月18日；



2017年1-2月刊 | 总第55期

人工智

封面报道 | Cover Report

## 规模会战下人工智能需要整合、转变

——对话上海交通大学电信学院认知机器和健康计算研究中心徐雷、连勇、杨杰、杨小康、张丽清教授

文本刊记者 杨焱

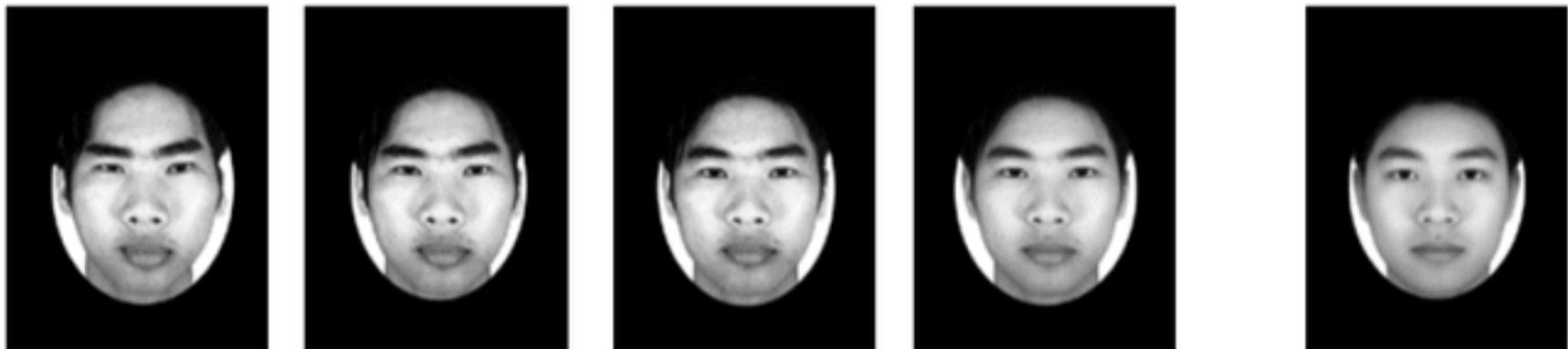
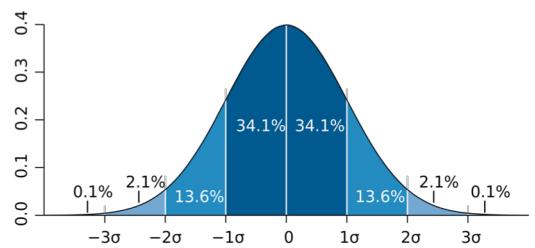


任重而道远！

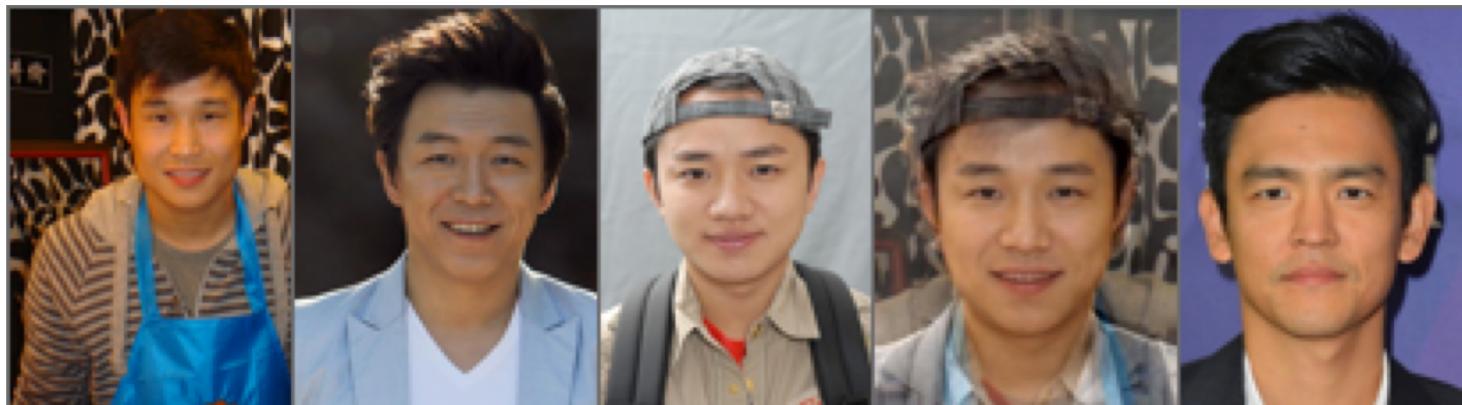
# 课程内容

- 聚类分析
  - 层次聚类、k-mean、GMM
- 最大似然学习、EM算法
- 线性模型（PCA，ICA，FA）
- 学习理论与模型选择
  - Bias-Variance, AIC, BIC, RPCL, Bayesian方法
- 因果发现
- 监督学习（线性回归、SVM）
- 神经网络（感知机、BP算法）
- 双向深度学习
  - AE，VAE，LMSER，KL对偶匹配

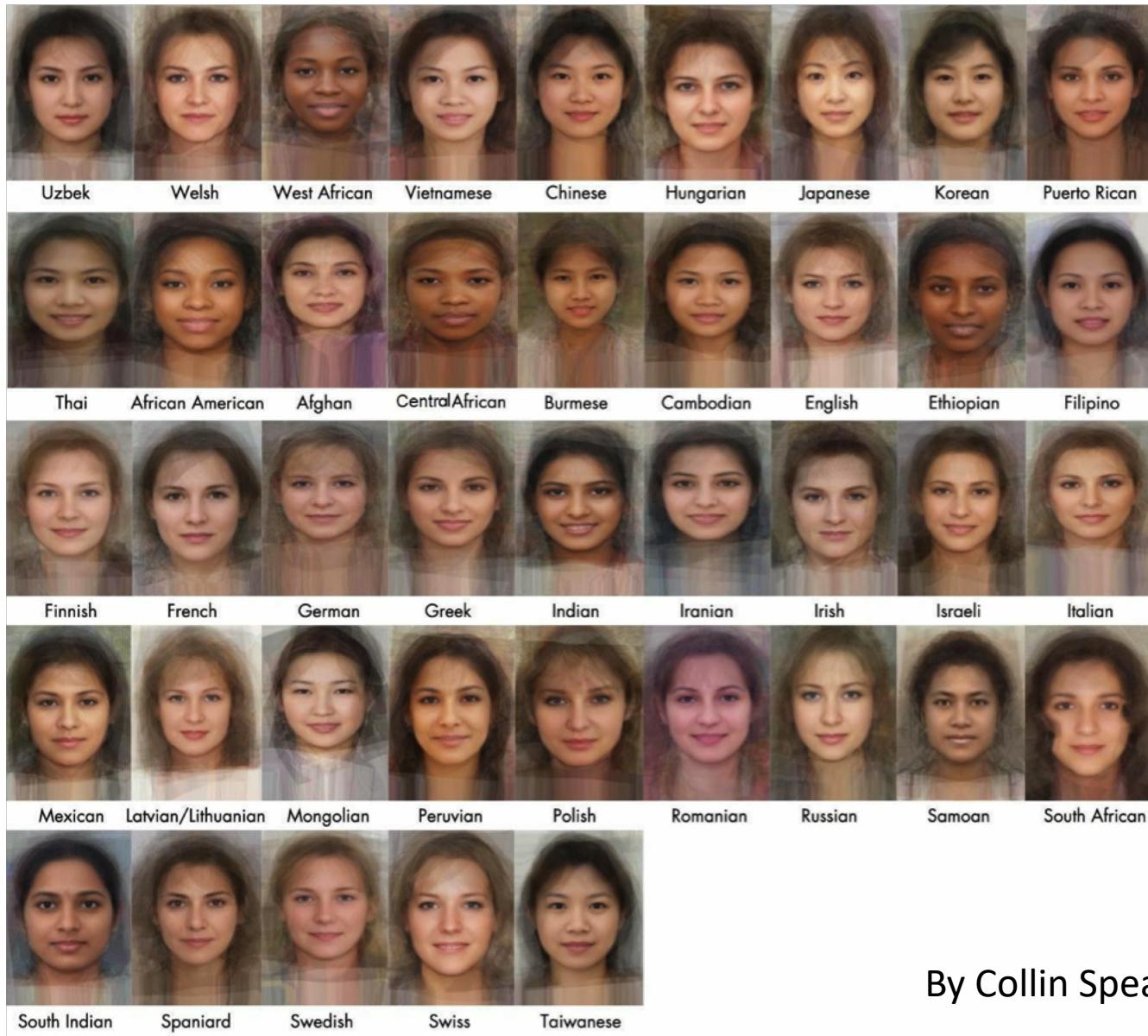
# 平均脸就是“好脸”



罗兹在论文中所使用的材料之一。最右一张是24名中国男性的平均脸，被用作标准模型。左一到左四的四张照片平均程度依次递增。在你看来，这四张脸的颜值是不是也依次递增呢？图片来源：参考文献[6]



# 各国女性的平均脸



By Collin Spears

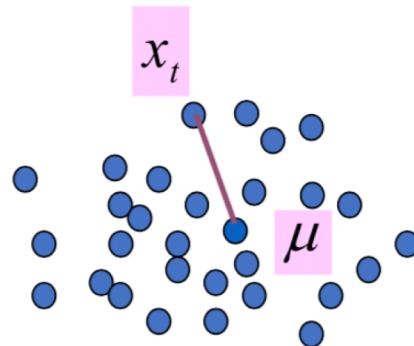
# 平方误差(MSE)与高斯分布

$$e_t = x_t - \mu, \quad (x_t - \mu)^2 \Leftrightarrow p(x_t, \theta) \propto e^{-c(x_t - \mu)^2}, c > 0 \text{ Gaussian}$$

$$G(x | \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x - \mu)^2}{2\sigma^2}\right\}$$

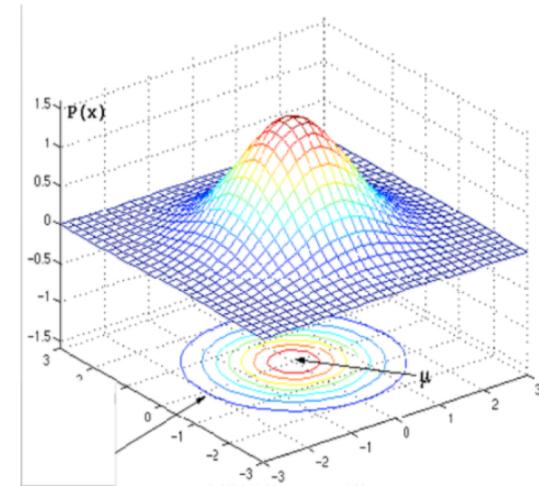
where  $\sigma^2$  is the variance, and  $\mu$  is the mean value.

$$e_t = x_t - \mu, \quad \|x_t - \mu\|^2 \Leftrightarrow p(x_t, \theta) \propto e^{-c\|x_t - \mu\|^2}, c > 0 \text{ Gaussian}$$



$$G(\mathbf{x} | \boldsymbol{\mu}, \sigma^2) = \frac{1}{(2\pi)^{d/2}\sigma^d} \exp\left\{-\frac{1}{2\sigma^2}(\mathbf{x} - \boldsymbol{\mu})^T(\mathbf{x} - \boldsymbol{\mu})\right\}$$

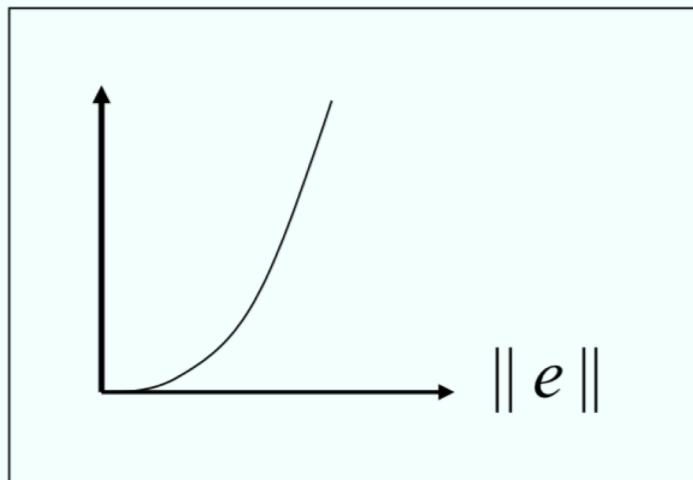
where  $\sigma^2$  is the variance, and  $\boldsymbol{\mu}$  is the mean vector.  
 $d$  is the dimension.



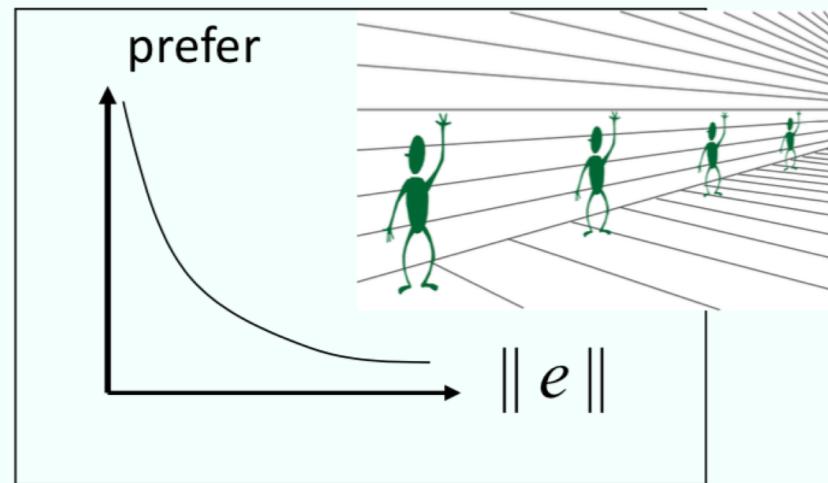
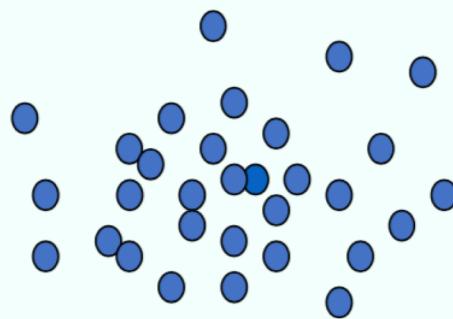
## Fitting cost versus likelihood

似然度

The farther distance, the smaller belonging chance (probability)

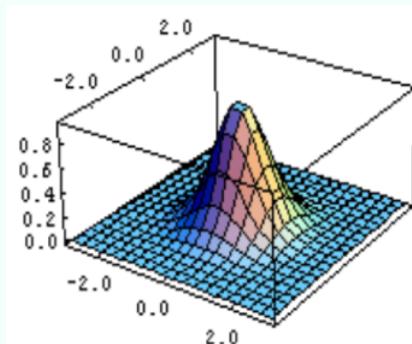


小错常犯,大错稀罕



$$q(e|\theta) = f(\|e\|)/C$$

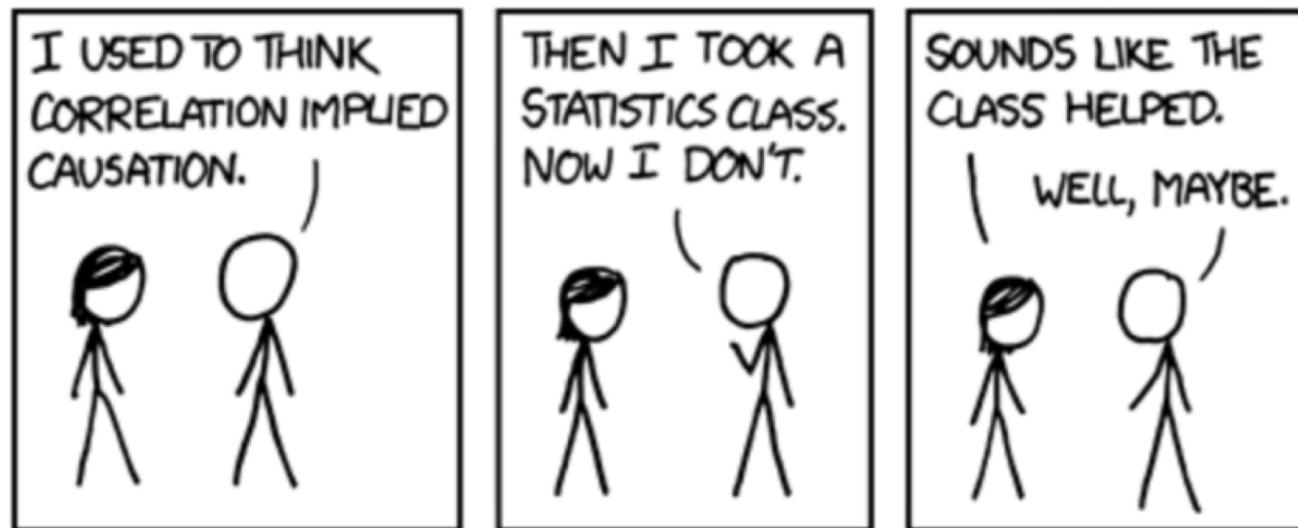
$$\int f(\|e\|)de = C < \infty$$



# 因果发现

## Causality vs. Dependence

- Causality → dependence ! Dependence → causality



(<http://imgs.xkcd.com/comics/correlation.png>)

X and Y are associated iff

$$\exists x_1 \neq x_2 P(Y|X=x_1) \neq P(Y|X=x_2)$$

X is a cause of Y iff

$$\exists x_1 \neq x_2 P(Y|\text{do}(X=x_1)) \neq P(Y|\text{do}(X=x_2))$$

29

# 因果发现：例子

The Telegraph

Home Video News **World** Sport Business Money Comment Culture Travel Life W...

USA | Asia | China | Europe | Middle East | Australasia | Africa | South America | Central Asia | N...

France | Francois Hollande | Germany | Angela Merkel | Russia | Vladimir Putin | Greece | Spain

HOME » NEWS » WORLD NEWS » EUROPE

Couples who share the housework are more likely to divorce, study finds

Divorce rates are far higher among “modern” couples who share the housework than in those where the woman does the lion’s share of the chores, a Norwegian study has found.



Couples may share the washing up but the same can not be said for financial decisions, the DWP has found Photo: ALAMY

# 因果发现：例子

The Telegraph

Home Video News World Sport Business Money Comment Culture Travel Life W  
USA | Asia | China | Europe | Middle East | Australasia | Africa | South America | Central Asia | I  
France | Francois Hollande | Germany | Angela Merkel | Russia | Vladimir Putin | Greece | Spain

HOME » NEWS » WORLD NEWS » EUROPE

Couples who share the housework are more likely to divorce, study finds

Divorce rates are far higher among “modern” couples who share the housework, a Norwegian study has found. The findings, published in the journal *Journal of Marriage and the Family*, suggest that women in couples where men do more housework are twice as likely to divorce as women in couples where men do less.

The Atlantic

Popular

Latest

Sections ▾

Magazine ▾

More ▾

Subscribe



GLOBAL

## Does Sharing Housework Really Lead to Divorce?

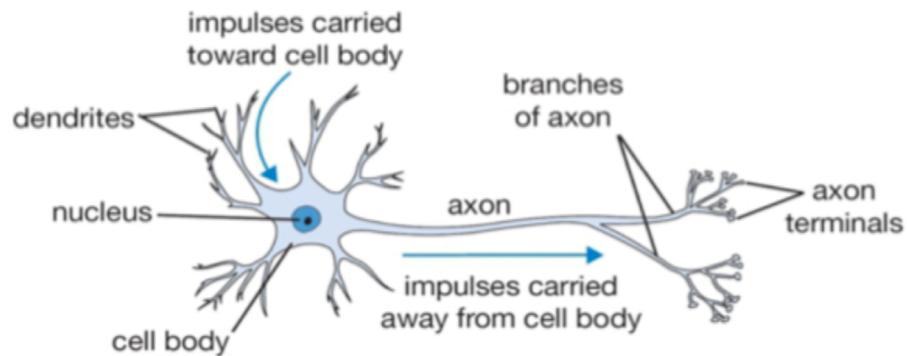
There's a study in the news that's bound to get a bunch of people talking ([Drudge tweeted it this morning](#), for instance, with more than 100 retweets). Whether those people are for or against its pronouncements, it seems to fly in the face of what we thought we knew about marriage, gender equality, and the way modern, successful relationships work.

JEN DOLL SEP 28, 2012

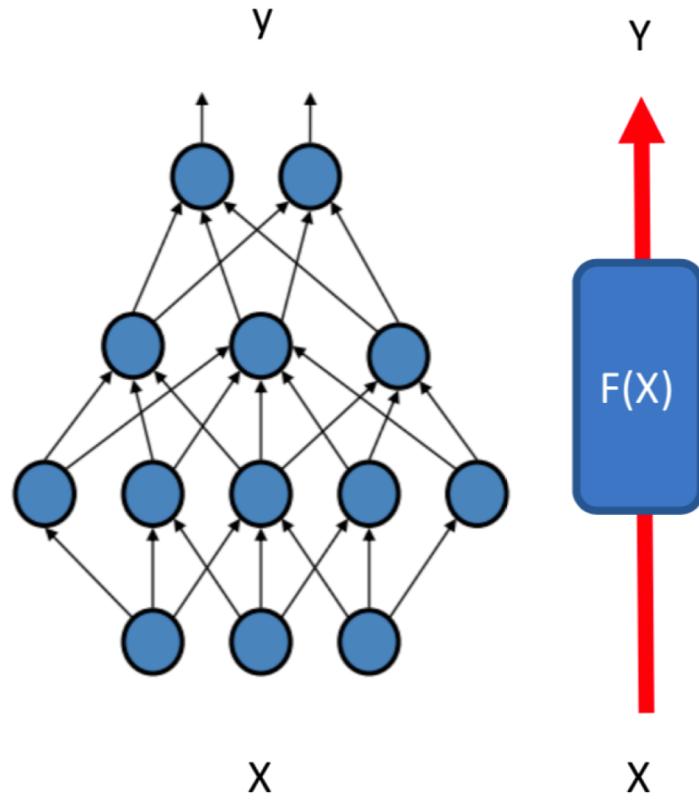
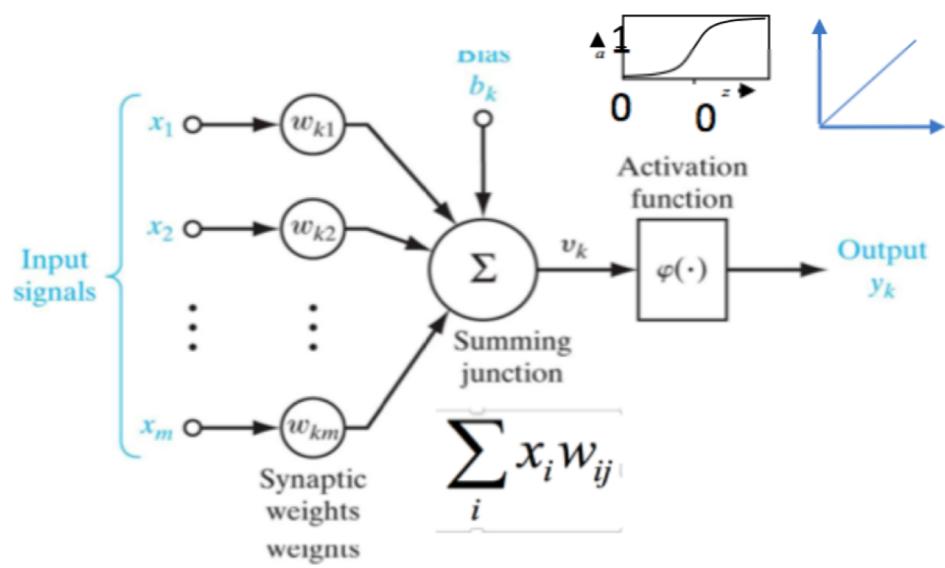


Couples may share the washing up but the same can not be said for financial decisions, the DWP has found Photo: ALAMY

# 简化的神经元模型 (1943年)

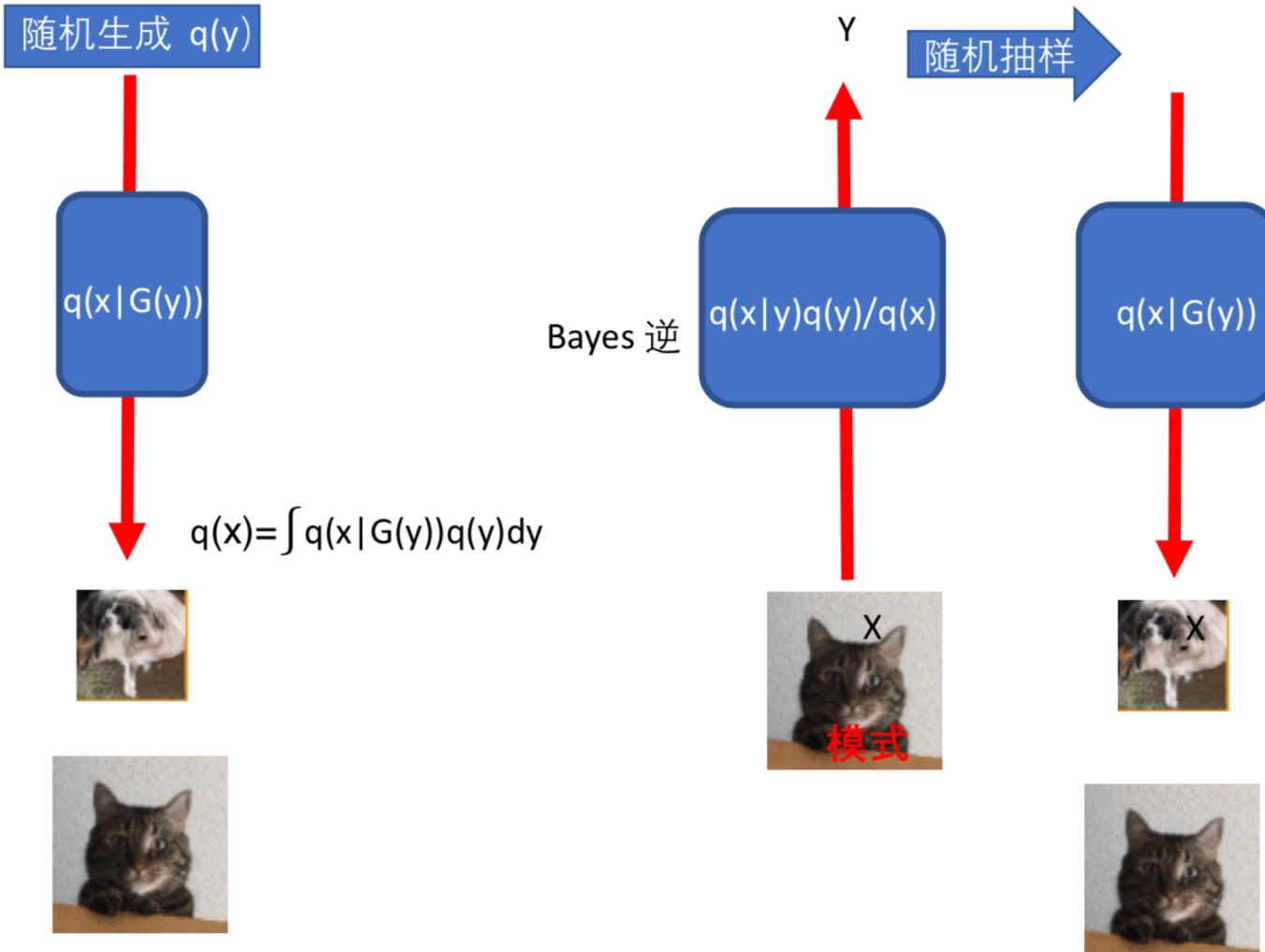


心理学家W.S.McCulloch  
数理逻辑学家W.Pitts



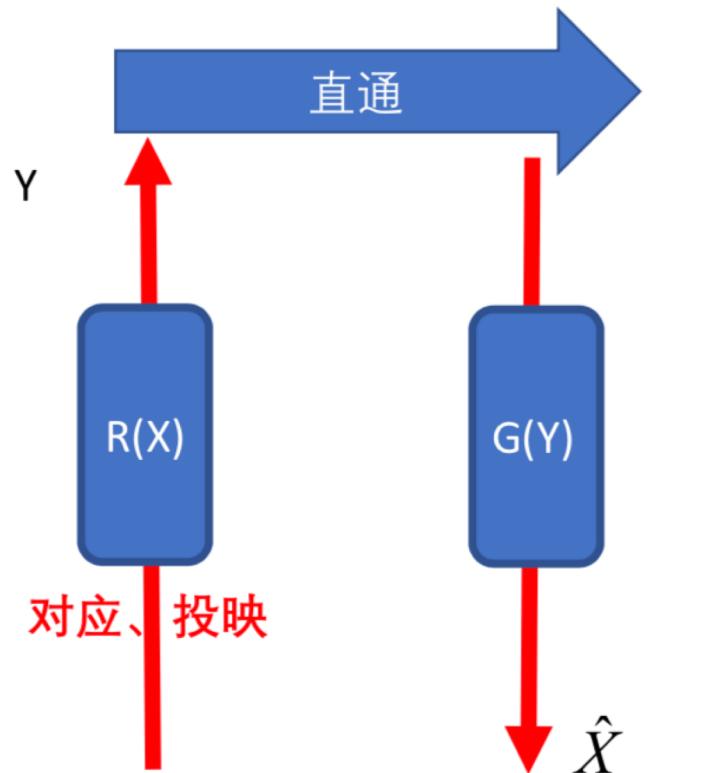
# 双向深度学习

## Maximum Likelihood Learning



# Autoencoder

标签

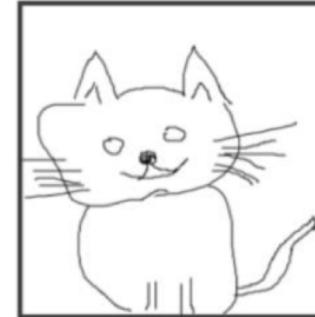
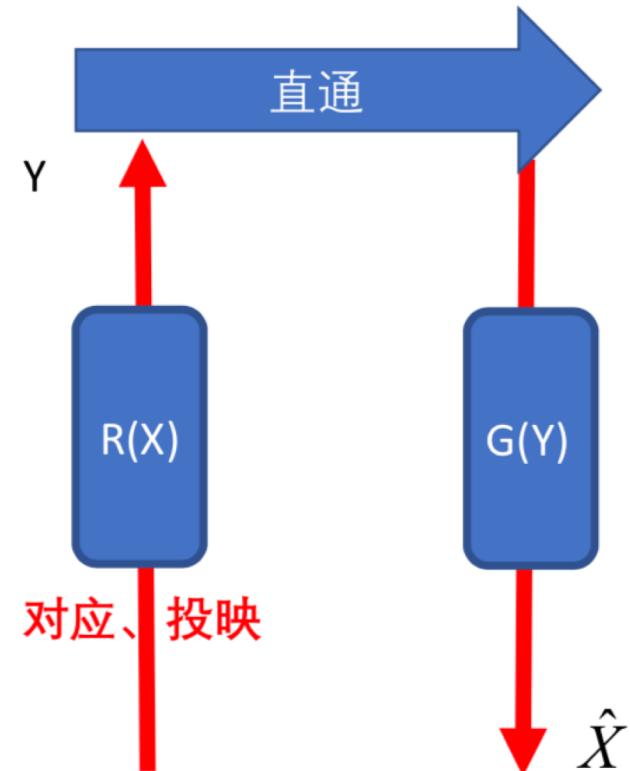


$X$

图像生成

$X$

标签

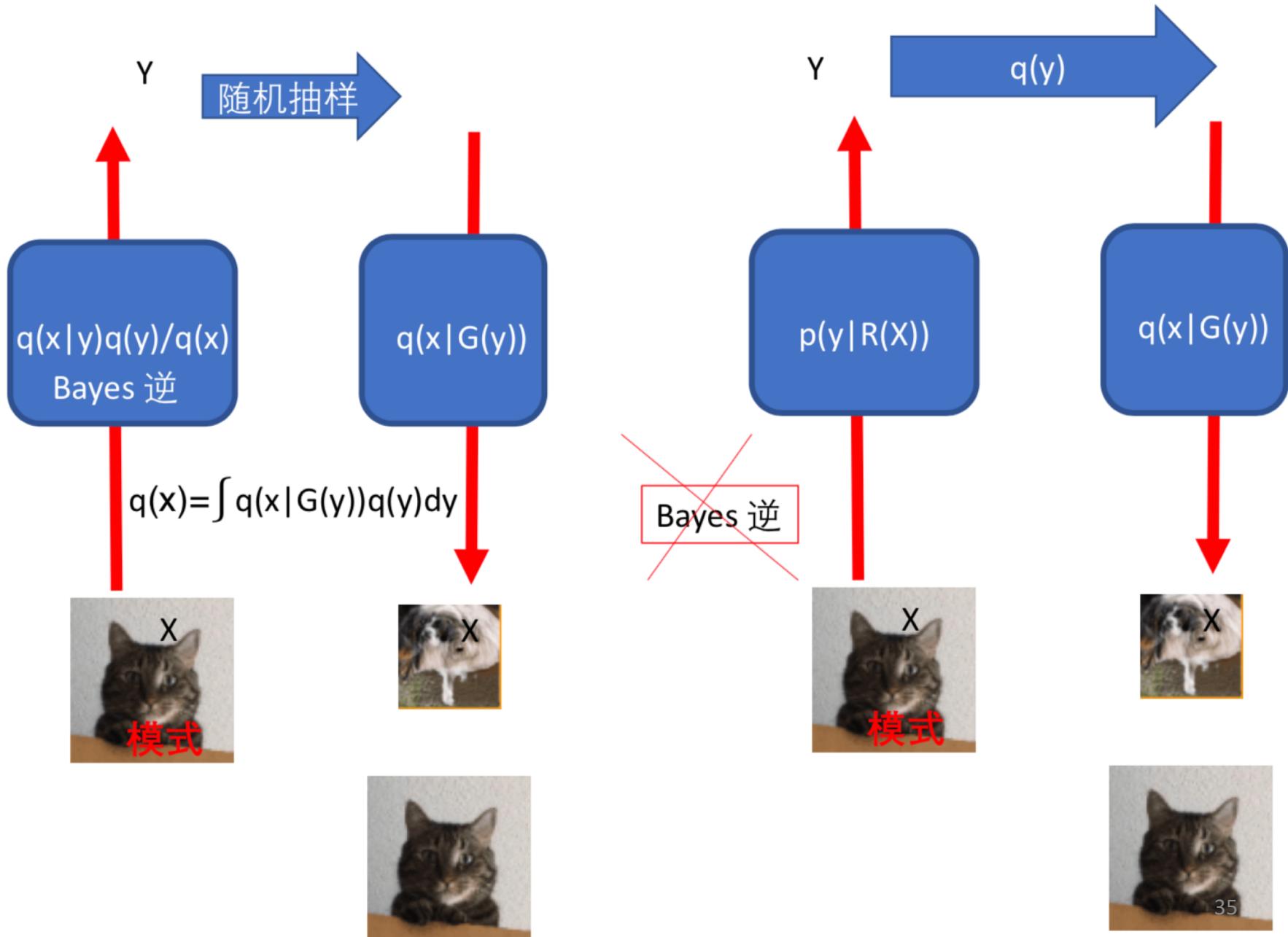


$Z$

条件图像生成

$X$  34

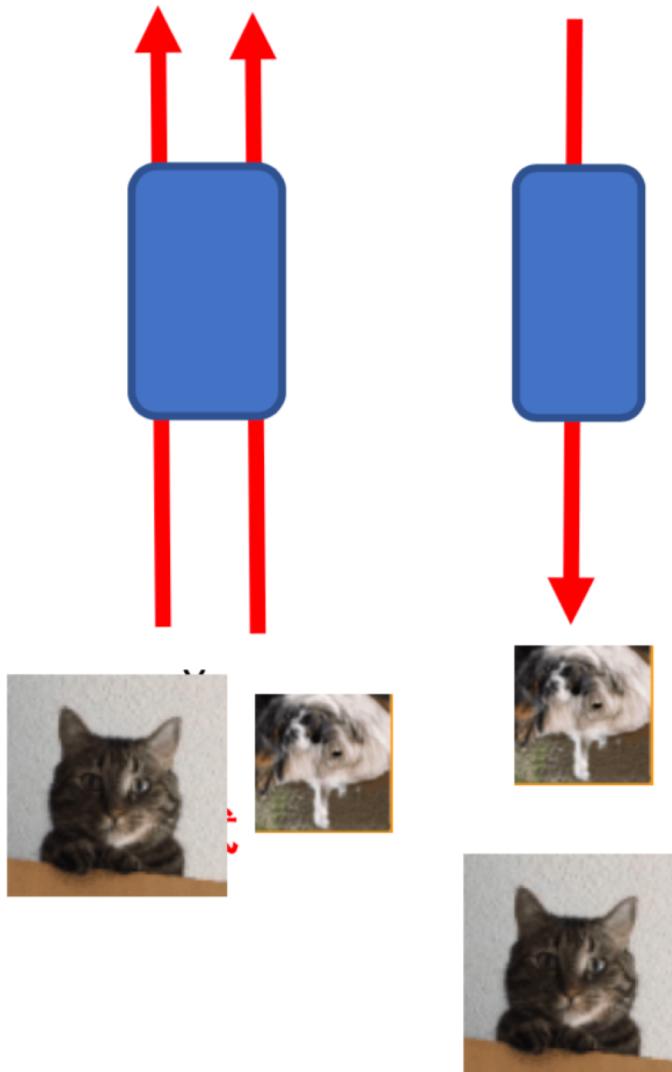
# Variational Autoencoder (VAE)



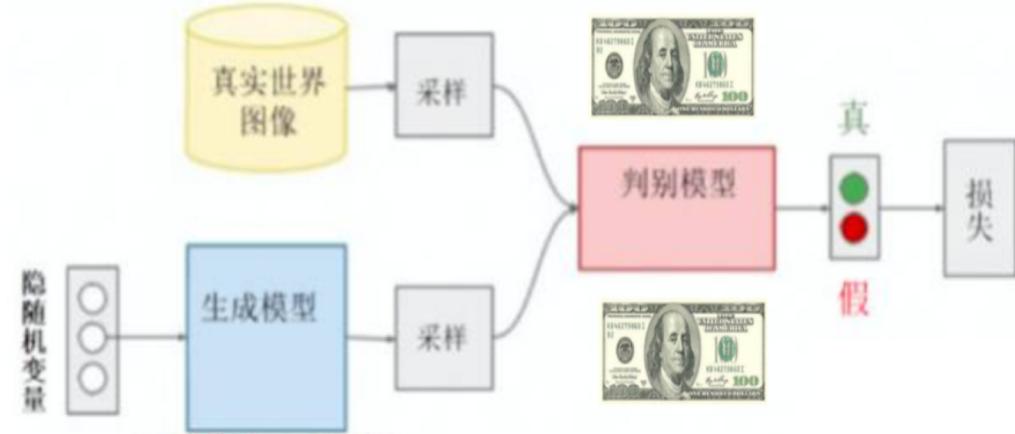
# 对抗学习 ?

Y

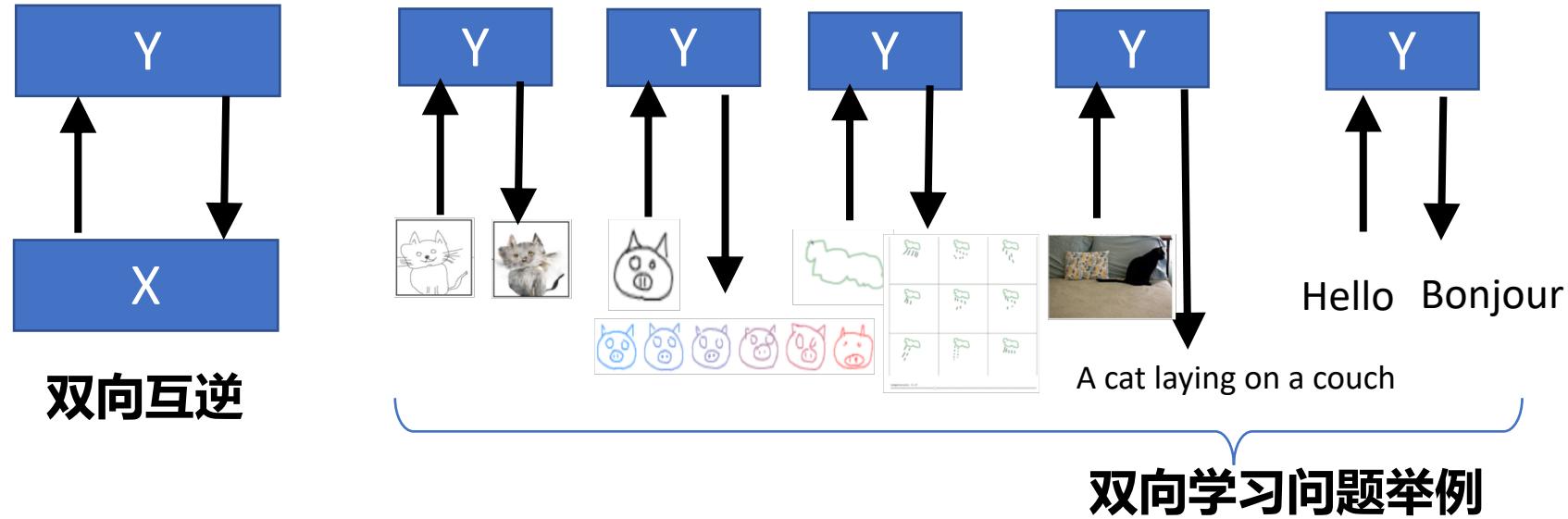
随机生成



generative adversarial net (GAN)

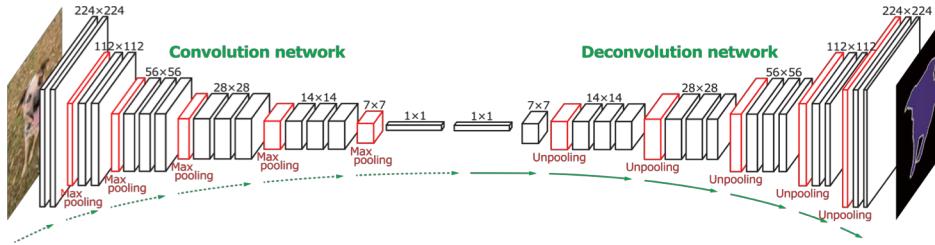


# 双向学习

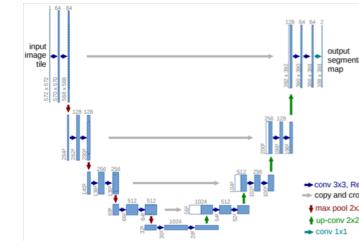


## 目前典型的双向深度网络结构

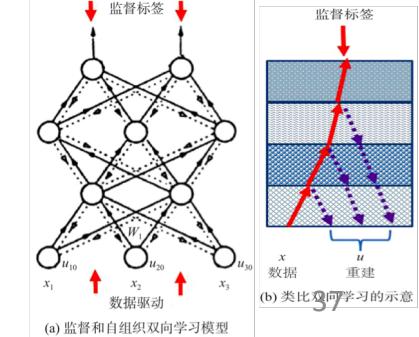
### • 卷积与反卷积



### • U-net结构



### • LMSER



(a) 监督和自组织双向学习模型

# **Thank you!**