

经典模型

颜少华

办公地点：C303，A501

办公时间：8：00 - 17：30

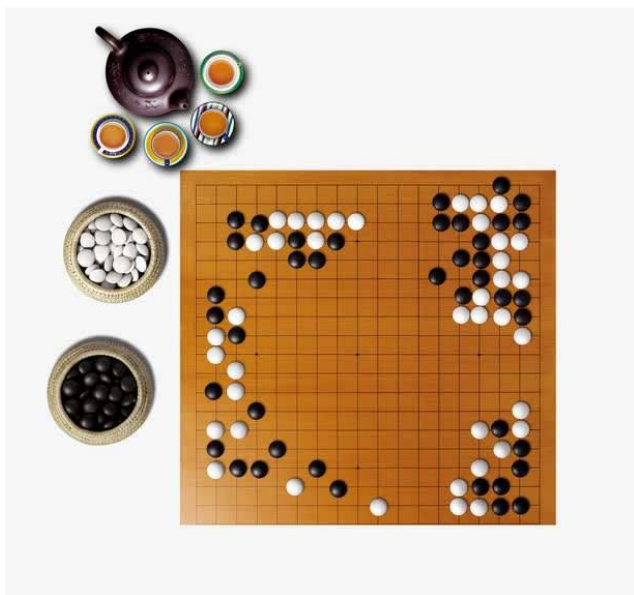
电话：17395282136

课程介绍

- 课时：80
- 上课时间：8：00 - 9：30；9：45 - 11：15
- 课程内容：逻辑回归和神经网络基础、图聚类、AlphaGo Zero、自然语言处理入门(时间有限可能不讲)。
- 成绩组成：出勤(10)+作业(30)+考试(60)(待定)

学完这门课能干什么

- 本课程侧重理论学习，示例所用数据集均为演示所用，数据规模小，但是本课程的算法会在《机器学习框架》课程中调用各类库进行实现，并使用更加接近实际的数据集。



部分数学家生卒年(排名不分先后)

数学家	生卒年	数学家	生卒年
牛顿	1643-1727	莱布尼茨	1646-1716
高斯	1777-1855	贝叶斯	1702-1761
欧拉	1707-1783	拉格朗日	1736-1813
拉普拉斯	1749-1827	卡尔达诺	1501-1576
凯利	1821-1895	阿基米德	前287年-前212
欧几里得	前330年—前275	闵可夫斯基	1864—1909
马尔可夫	1856-1922	刘徽	约225-约295
祖冲之	429-500	祖暅	456-536

经典

核心价值不会随时间流逝而改变的事物，被称作为经典。

经久不衰的万世之作，后人尊敬它称之为经典。

经典是指具有典范性、权威性的著作。

经典就是经过历史选择出来的“最有价值的书”。

classical ☆ ⋮

英 /ˈklæsɪk(ə)/



美 /ˈklæsɪkəl/



全球(英国)



简明

牛津

新牛津 NEW

朗文

韦氏 NEW

柯林斯

例句

adj. 古典的；经典的；传统的；第一流的

n. 古典音乐

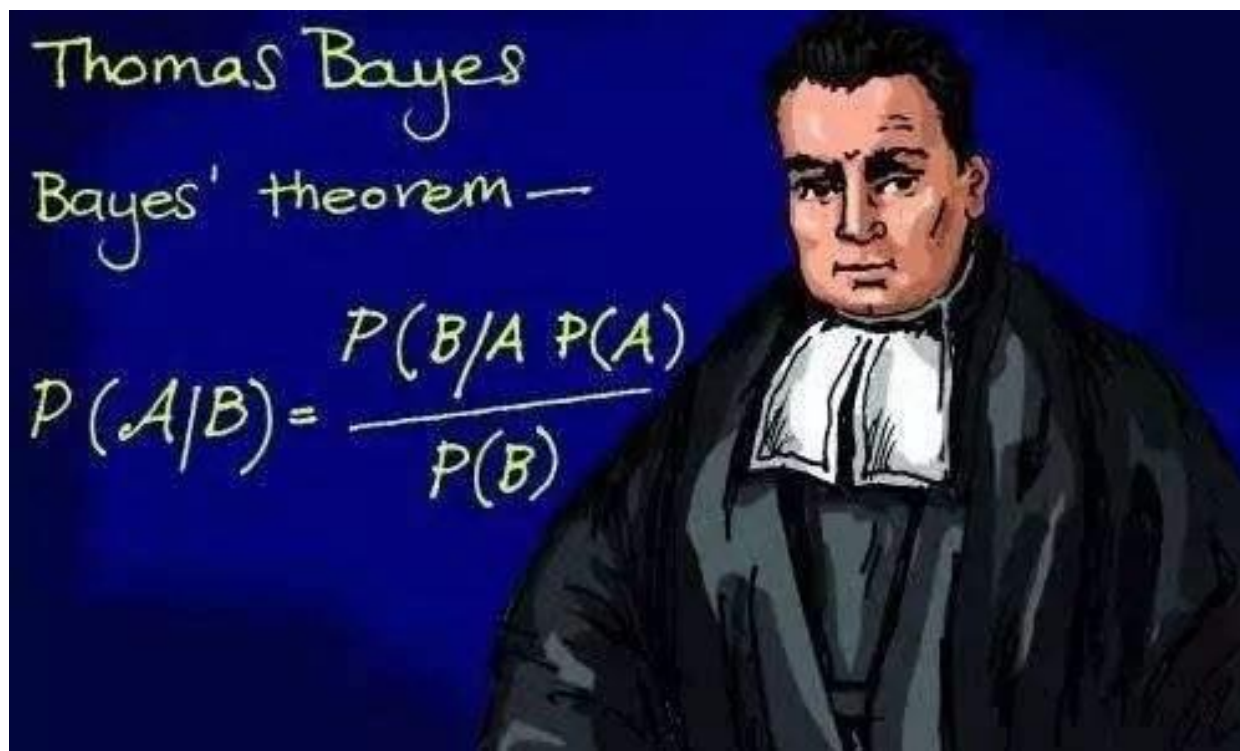
人工智能与其他学科

- 世界上第一台通用计算机“ENIAC”于1946年2月14日在美国宾夕法尼亚大学诞生。
- 人工智能(AI)这一称呼起源于1956年的达特茅斯会议。
- 但人工智能的理论基础早就已经根植于其他学科中。

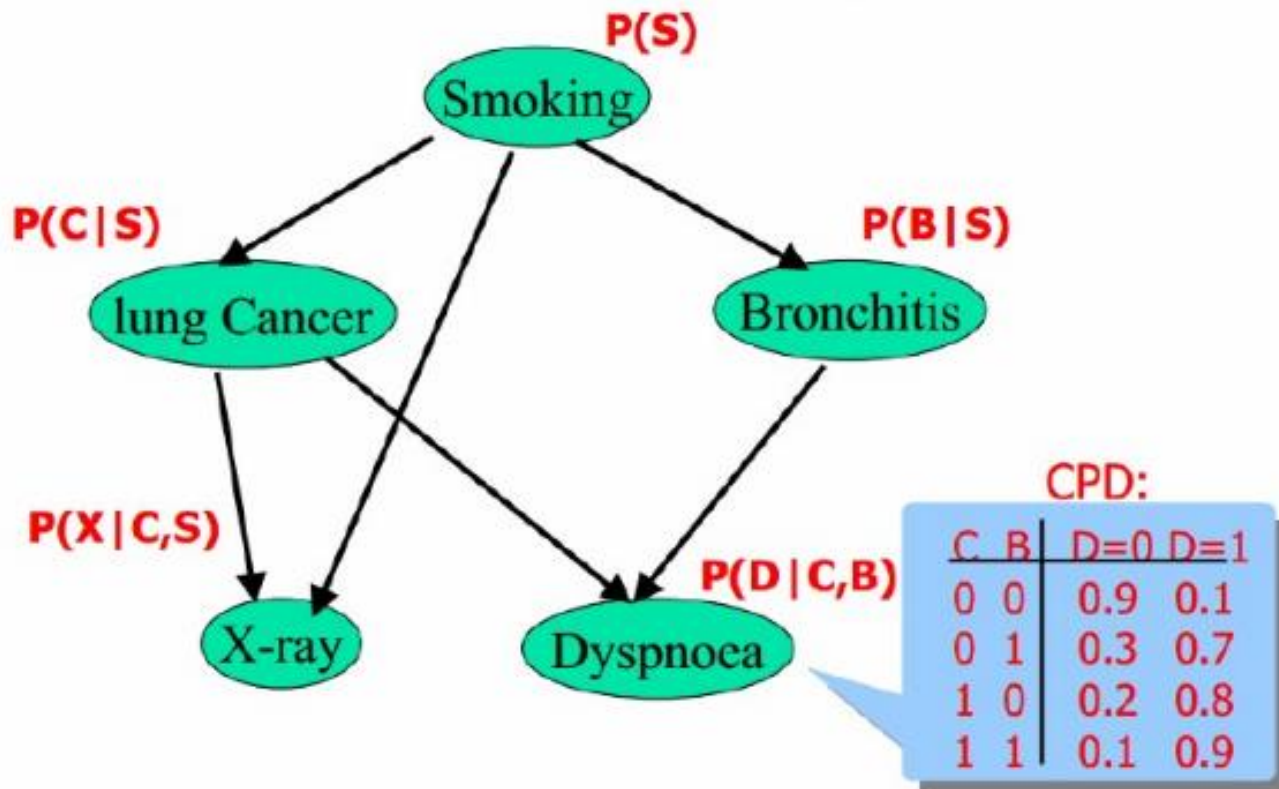


朴素贝叶斯

贝叶斯(约1702-1761)首先将归纳推理法用于概率论基础理论，并创立了贝叶斯统计理论。



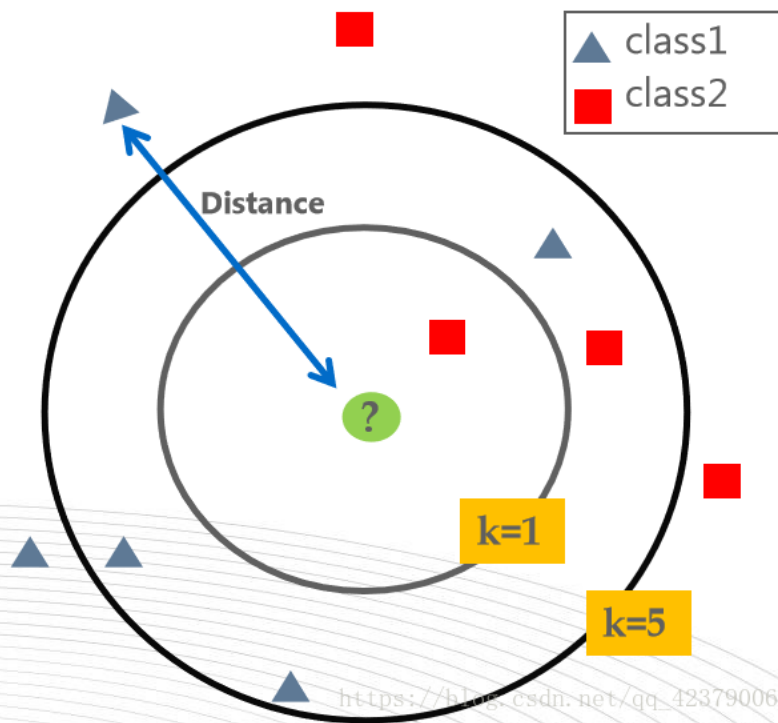
贝叶斯网络



K近邻

KNN (K Nearest Neighbor) :

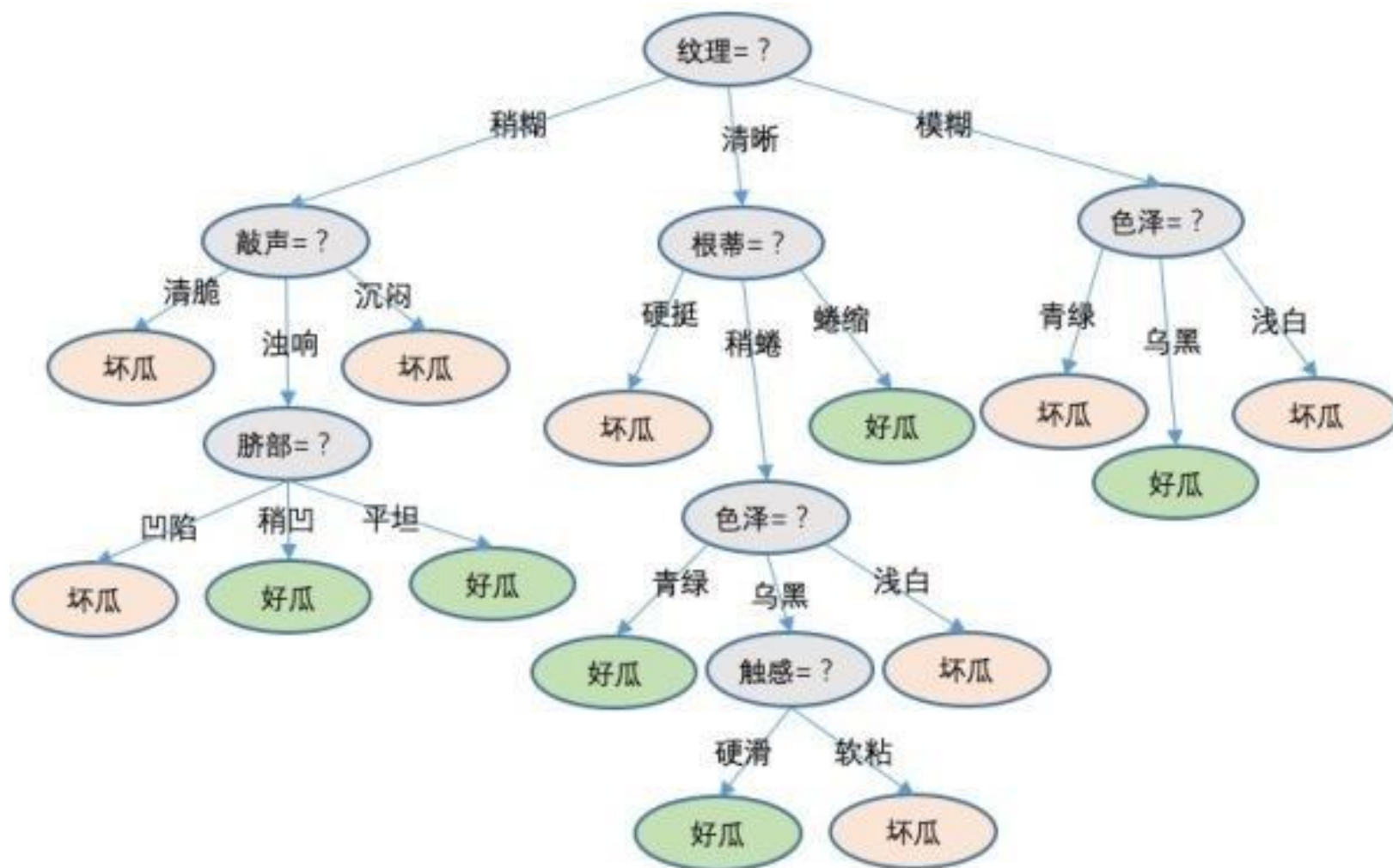
k个最近的邻居，即每个样本都可以用它最接近的k个邻居来代表。



K近邻

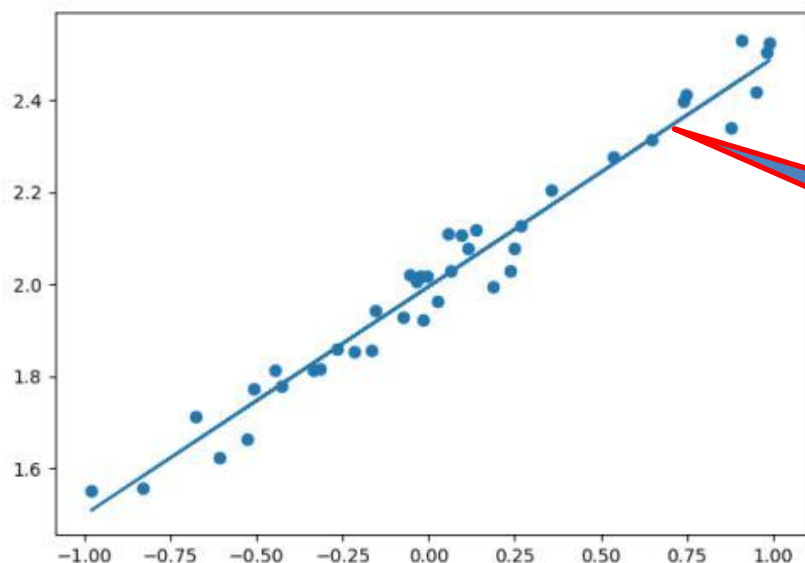


决策树，随机森林

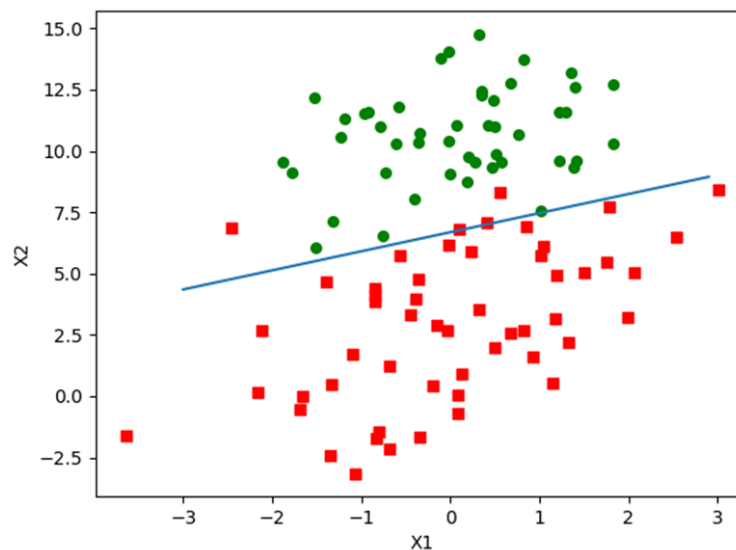


基于表生成含有缺失值的决策树

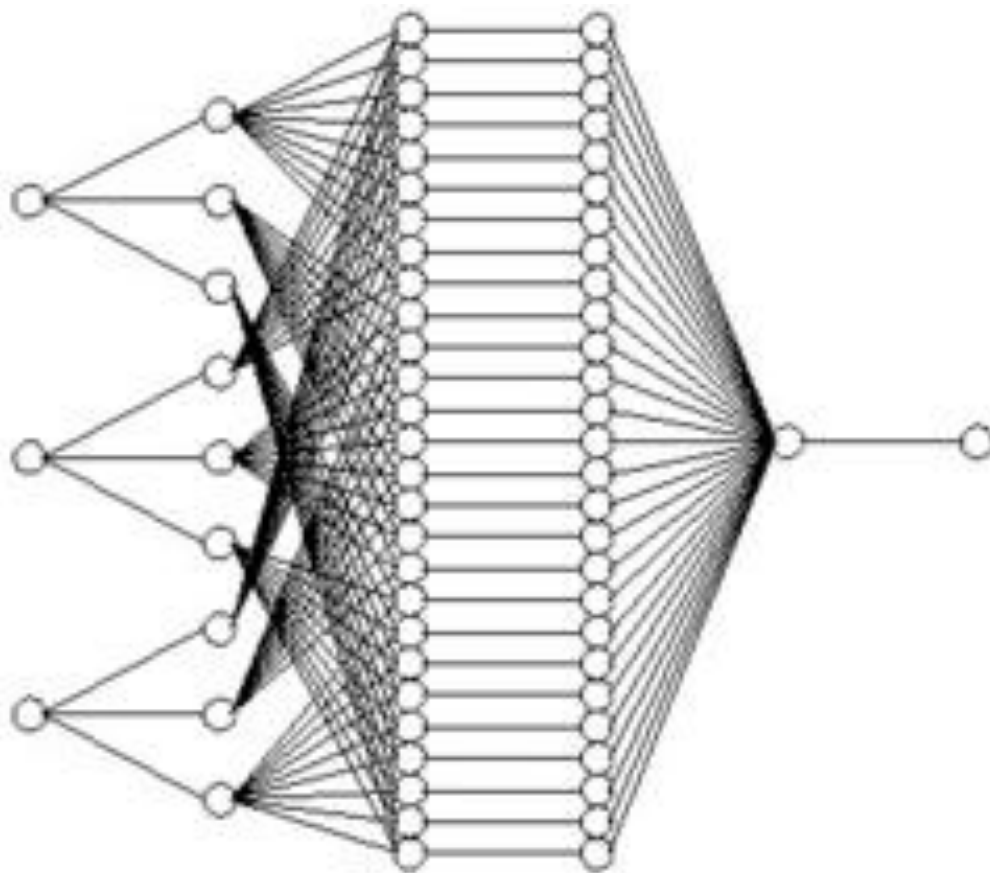
线性回归与逻辑回归



这条线似乎把数据分成了两部分，那能不能像下面的图那样用来做分类呢。

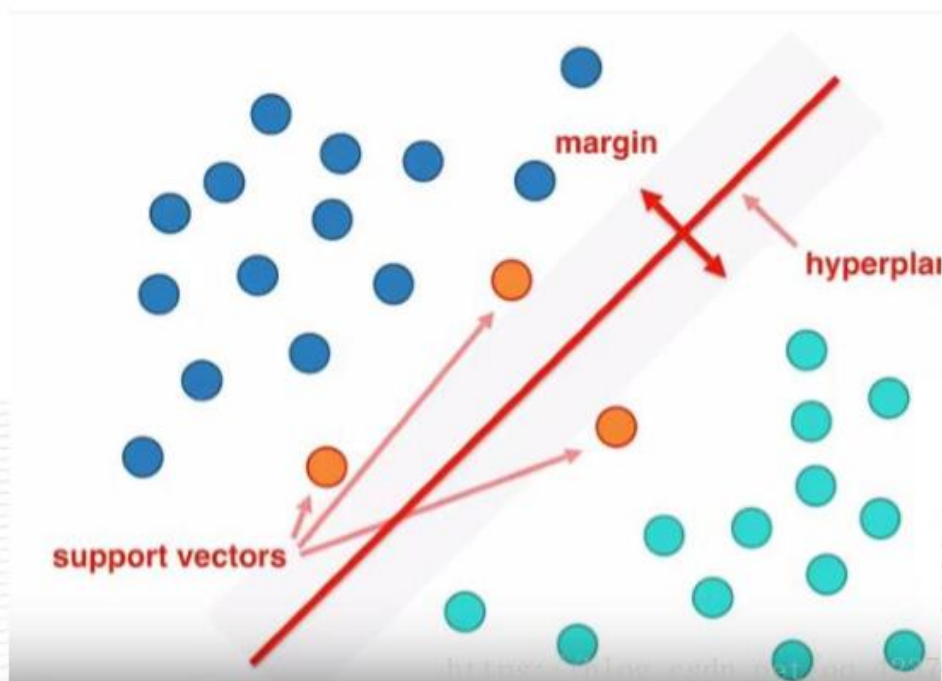


神经网络基础



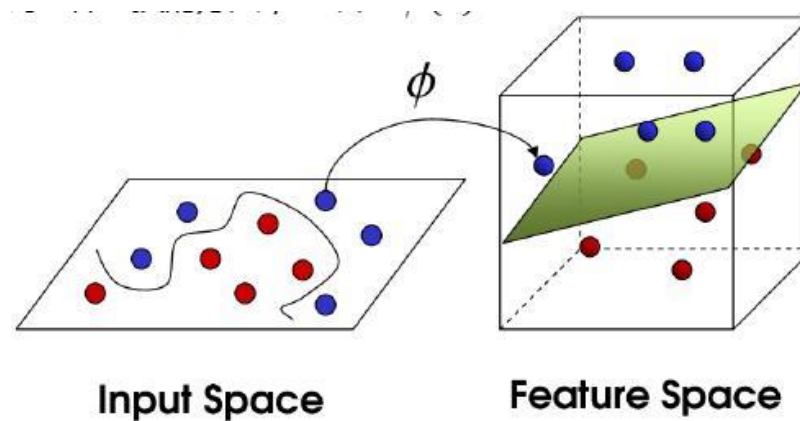
支持向量机(Support Vector Machine, SVM)

SVM：寻找到一个超平面使样本分成两类，并且间隔最大。而我们求得的 w 就代表着我们需要寻找的超平面的系数。



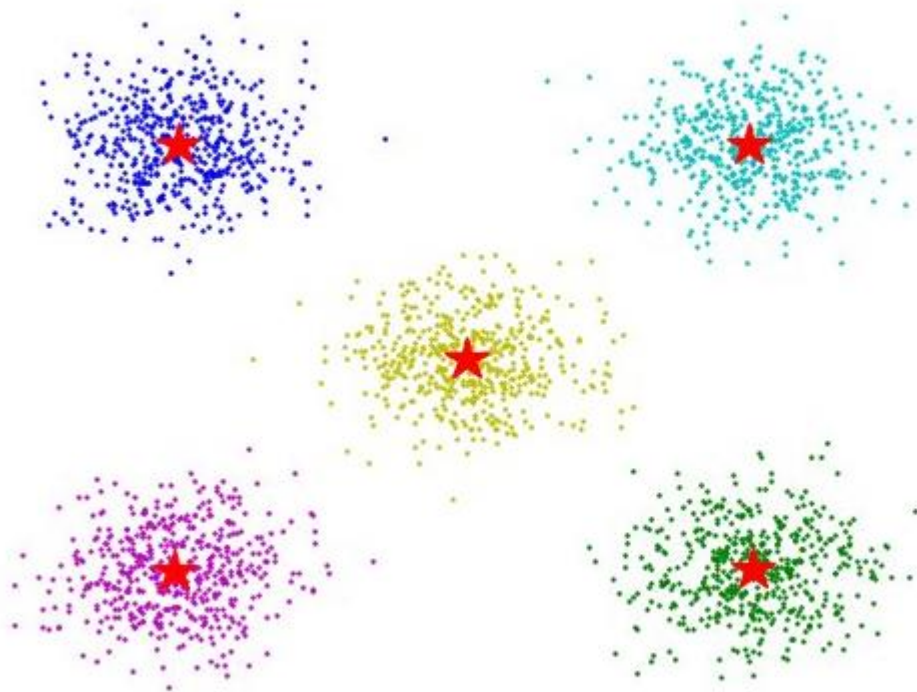
支持向量机

- 核函数

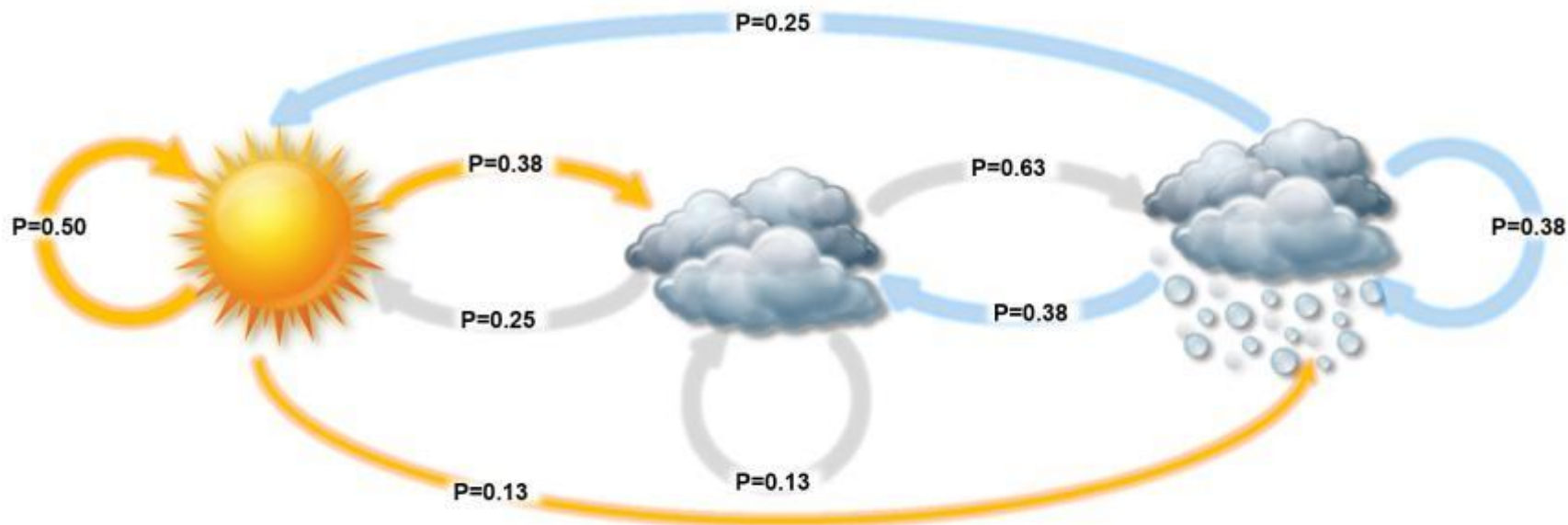


K均值

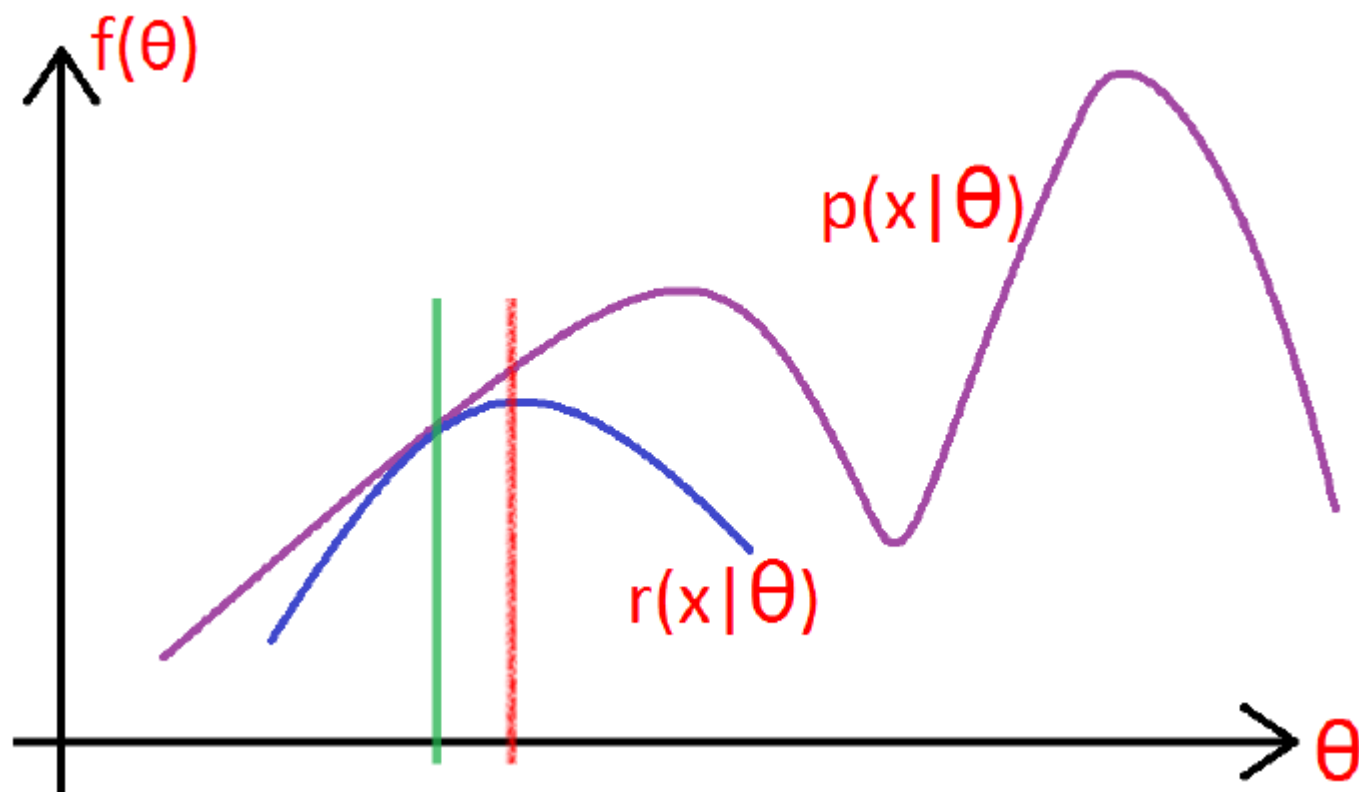
KMeans Clustering



马尔可夫



EM算法



EM算法

$$\gamma(i, k) = \frac{\pi_k N(x_i | \mu_k, \Sigma_k)}{\sum_{j=1}^K \pi_j N(x_i | \mu_j, \Sigma_j)}$$

$$\begin{cases} \mu = \frac{1}{n} \sum_i x_i \\ \sigma^2 = \frac{1}{n} \sum_i (x_i - \mu)^2 \end{cases}$$

$$\begin{cases} N_k = \sum_{i=1}^N \gamma(i, k) \\ \mu_k = \frac{1}{N_k} \sum_{i=1}^N \gamma(i, k) x_i \\ \Sigma_k = \frac{1}{N_k} \sum_{i=1}^N \gamma(i, k) (x_i - \mu_k)(x_i - \mu_k)^T \\ \pi_k = \frac{N_k}{N} = \frac{1}{N} \sum_{i=1}^N \gamma(i, k) \end{cases}$$

特殊的数据集

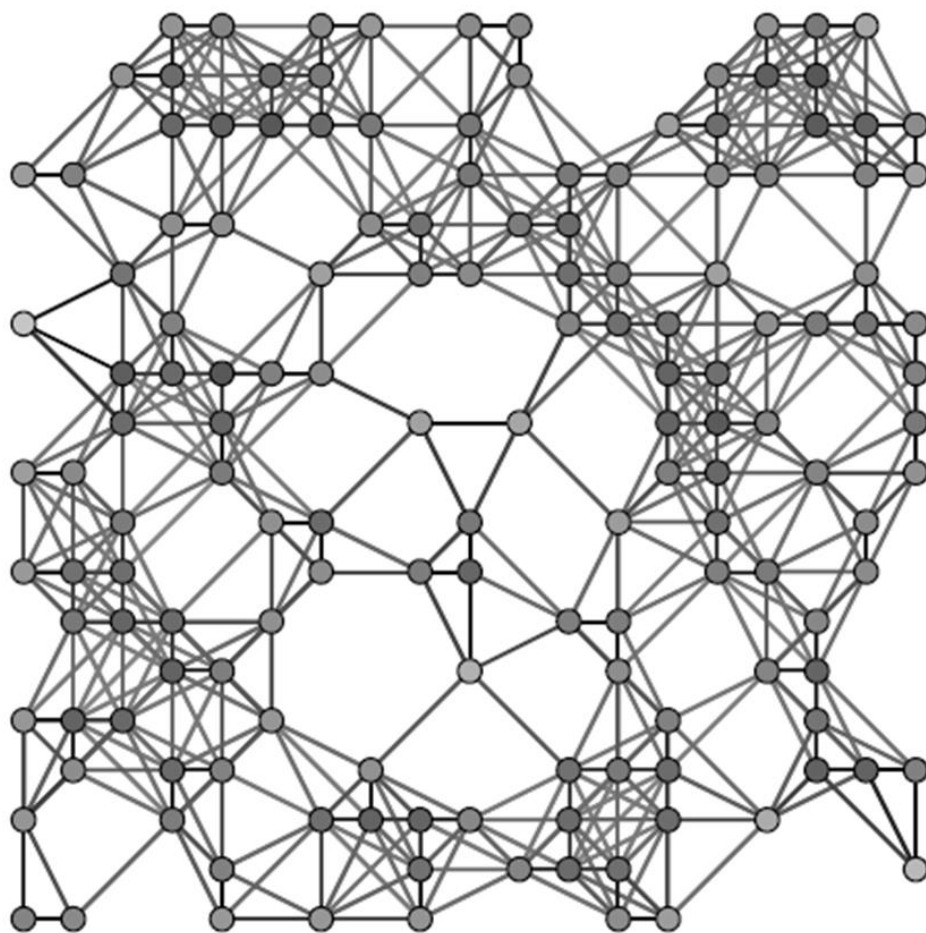
A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
表1 总人口、户籍人口、少数民族人口比重、非农业户口人口比重、城乡人口、家庭户人口、家庭户类别（全部数据）																	
县市区	总人口（人）			总人口性别比	户籍人口（人）	少数民族人口比重（%）	非农业户口人口比重（%）	城乡人口（人）		家庭户				家庭户类别（户）			
	合计	男	女					城镇	乡村	户数（户）	人口数（人）	规模（人/户）	其中：一人户（户）	一代户	二代户	三代户	四代以上户
北京市	13569194	7074518	6494676	108.93	11167502	4.31	60.17	0	3046730	4096844	11922945	2.91	494553	1267259	2240735	578024	10826
市辖区	11509595	6020903	5488692	109.70	9285488	4.35	66.20	10522464	1632934	3483627	9983850	2.87	443010	1105041	1892079	478676	7831
东城区	525558	260407	266066	101.70	625381	6.40	80.72	0876661		160772	480705	2.88	72877	50061	00787	78505	780

以前课上的数据集基本都是向量形式的，
那么如果是这种形式的怎么办呢？

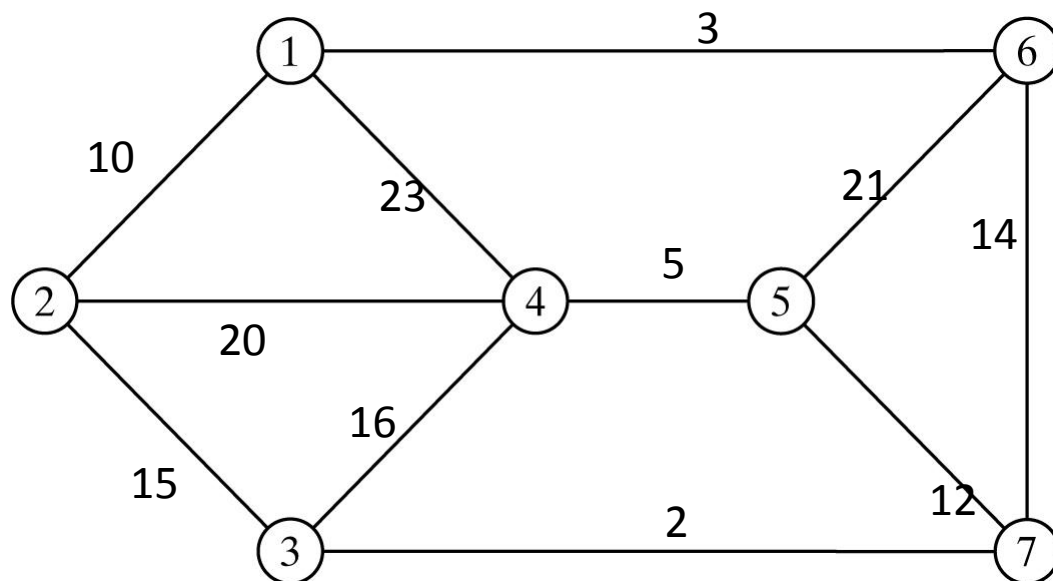
```
4.8, 3.0, 1.4, 0.1, Iris-setosa
5.7, 4.4, 1.5, 0.4, Iris-setosa
5.7, 3.8, 1.7, 0.3, Iris-setosa
5.1, 3.7, 1.5, 0.4, Iris-setosa
4.8, 3.4, 1.9, 0.2, Iris-setosa
5.2, 3.5, 1.5, 0.2, Iris-setosa
4.8, 3.1, 1.6, 0.2, Iris-setosa
5.5, 4.2, 1.4, 0.2, Iris-setosa
5.5, 3.5, 1.3, 0.2, Iris-setosa
5.1, 3.4, 1.5, 0.2, Iris-setosa
4.4, 3.2, 1.3, 0.2, Iris-setosa
4.8, 3.0, 1.4, 0.3, Iris-setosa
5.3, 3.7, 1.5, 0.2, Iris-setosa
6.4, 3.2, 1.5, 0.5, Iris-versicol
```

张三，是王五的好朋友，刚认识李四，对赵六很是反感。那么，对于张三，我们无法直接得出他的特征，但能知道他的活动圈。利用图聚类，可以将同一社交范围的人聚合到一起。

特殊的数据集



图聚类1: 谱聚类



图聚类2: 马尔可夫聚类

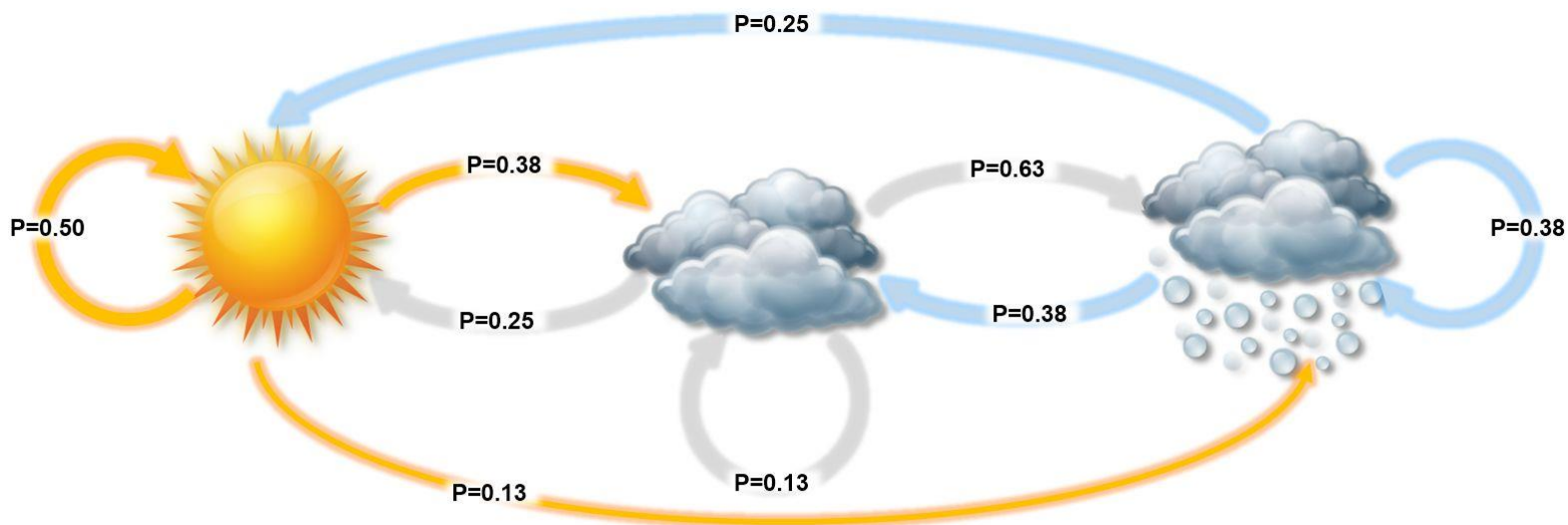


张松海

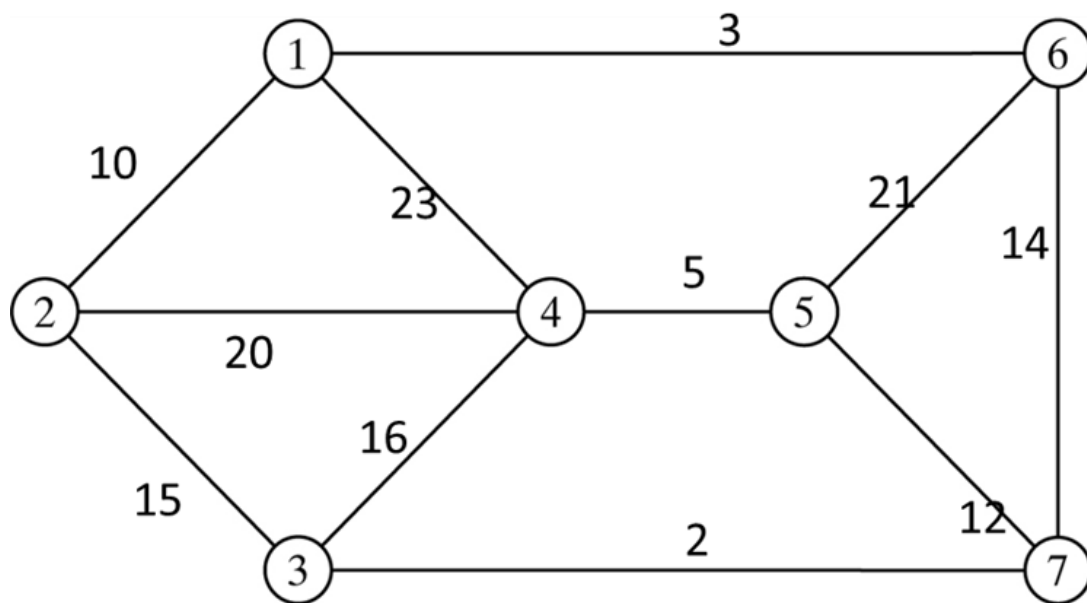
- 1 $P=NP?$
- 2 霍奇猜想
- 3 庞加莱猜想
- 4 黎曼假设

- 5 杨-米尔斯规范场
存在性和质量间
隔假设
- 6 NS方程解的存在
性与光滑性

- 7 贝赫和斯维讷通-
戴尔猜想



图聚类2：马尔可夫聚类

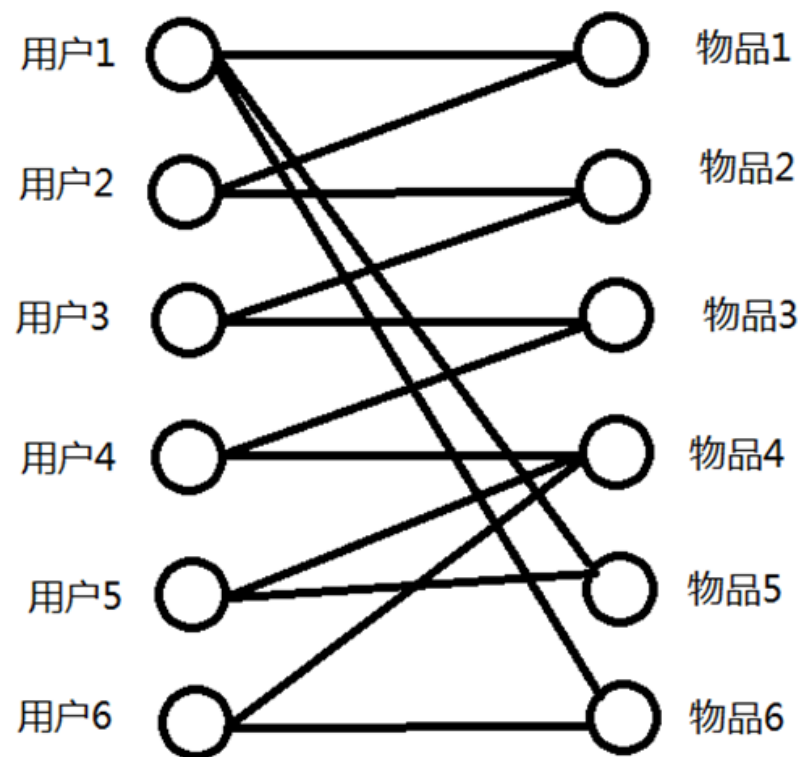


图的其他应用：推荐系统

	物品1	物品2	物品3	物品4
用户1	1	4	4	5
用户2	1	5	4	5
用户3	1	2	2	5
用户4	2	1	1	4

	物品1	物品2	物品3	物品4	物品5	物品6
用户1	1				1	1
用户2	1	1				
用户3		1	1			
用户4			1	1		
用户5				1	1	
用户6				1		1

图的其他应用：推荐系统



二分图

- 一个公司有 N 个岗位空缺，共有 M 人应聘，每个岗位需要有一定资格的人来填补，由于每个人的能力不同，所以不同的人能胜任不同的工作。
- 每个工作岗位只需要一个人，每个人也只能做一份工作。
- 现在已知每个人能胜任的工作，求这 M 个人最多能胜任几份工作。

二分图

◆ 假设 $N=5$ ， $M=5$ ，且每个人能胜任的工作如表中所示。

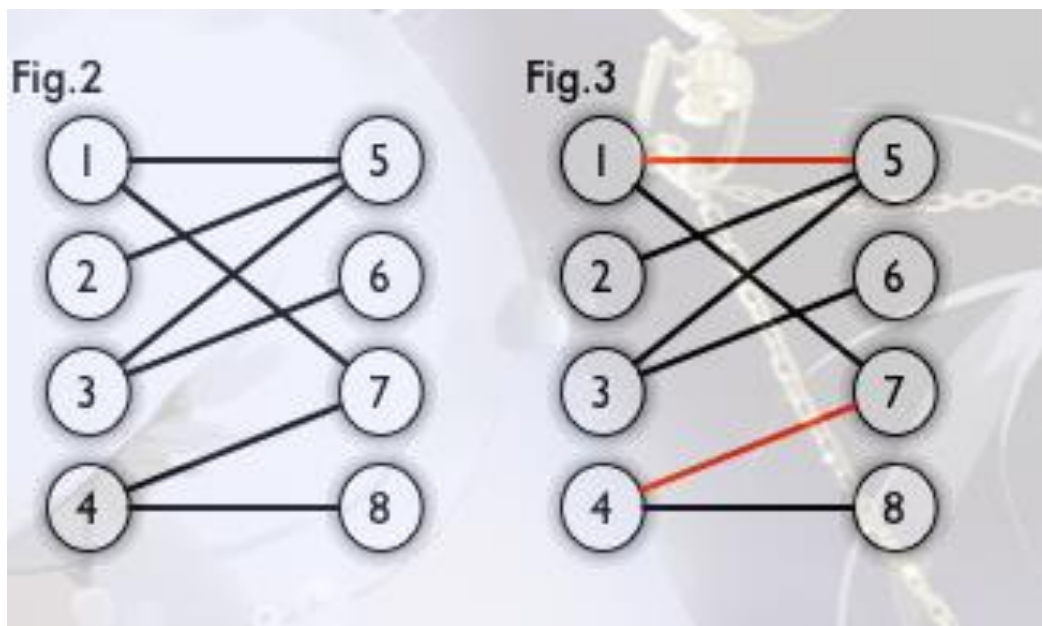
◆ 方案：

- ◆ 1号应聘者参加工作5
- ◆ 2号应聘者参加工作3
- ◆ 3号应聘者参加工作1
- ◆ 5号应聘者参加工作2

应聘者	能胜任的工作
1	2, 5
2	2, 3, 4
3	1, 5
4	1, 2, 5
5	2

二分图

- 在图 G 的一个子图 M 中， M 的边集中的任意两条边都不依附于同一个顶点，则称 M 是一个匹配。选择这样的边数最大的子集称为图的最大匹配问题，最大匹配的边数称为最大匹配数。

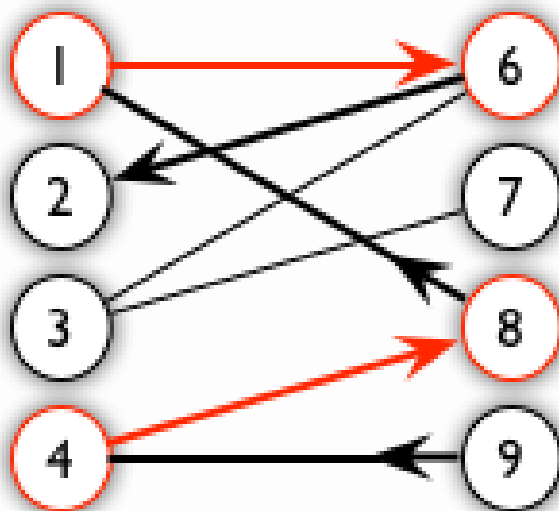


此图不是上面的招聘问题无关。

二分图

- 交替路：从一个未匹配点出发，依次经过非匹配边、匹配边、非匹配边...形成的路径叫交替路。

Fig.5



二分图

- 增广路：从一个未匹配点出发，走交替路，如果到达另一个未匹配点（出发的点不算），则这条交替路称为增广路（augmenting path）。例如，图 5 中的一条增广路如图 6 所示（图中的匹配点均用红色标出）：

Fig.5

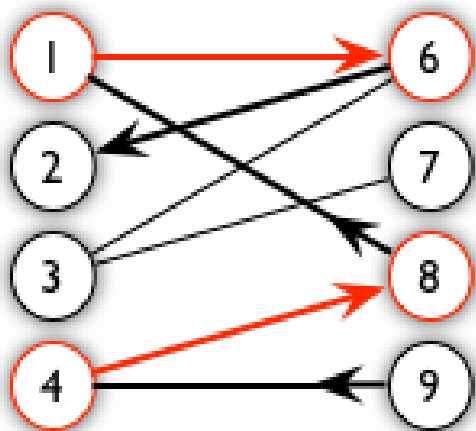
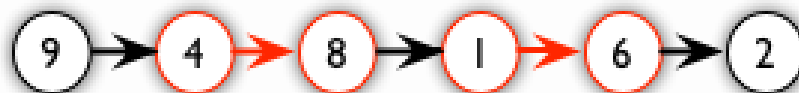


Fig.6



二分图

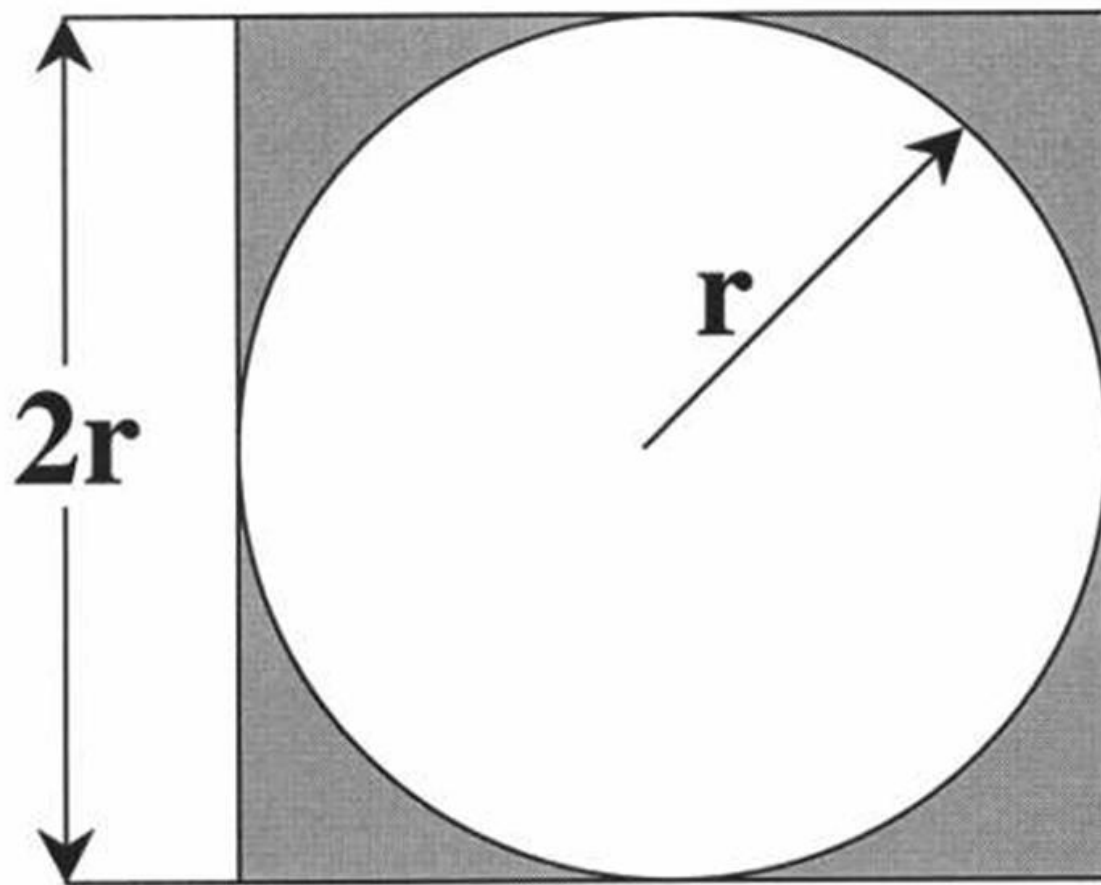
- 增广路有一个重要特点：非匹配边比匹配边多一条。因此，研究增广路的意义是改进匹配。只要把增广路中的匹配边和非匹配边的身份交换即可。由于中间的匹配节点不存在其他相连的匹配边，所以这样做不会破坏匹配的性质。交换后，图中的匹配边数目比原来多了 1 条。
- 我们可以通过不停地找增广路来增加匹配中的匹配边和匹配点。找不到增广路时，达到最大匹配(增广路定理)。

AlphaGo Zero

- AlphaGo Zero是谷歌下属公司Deepmind的新版程序。
- 从空白状态学起，在**无任何人类输入(规则需要人的输入)**的条件下，AlphaGo Zero能够迅速**自学**围棋，并以100:0的战绩击败“前辈”。



蒙特卡罗方法



理想与现实



🔍 搜狐网 - 百度快照

超越人类(人工智能)影评

2019年5月3日 - 超越人类 Mr.Chen 评论人工智能 4 2019-05-03 12:25:23 这篇影评可能有剧透 AI,大多数人都不能逃避的现实选择,不单是兴趣勾忆着我们的渴望,还有前景...

<https://movie.douban.com/revie...> - 百度快照

未来人工智能将超越人类、取代人类?霍金先生生前表示担忧!

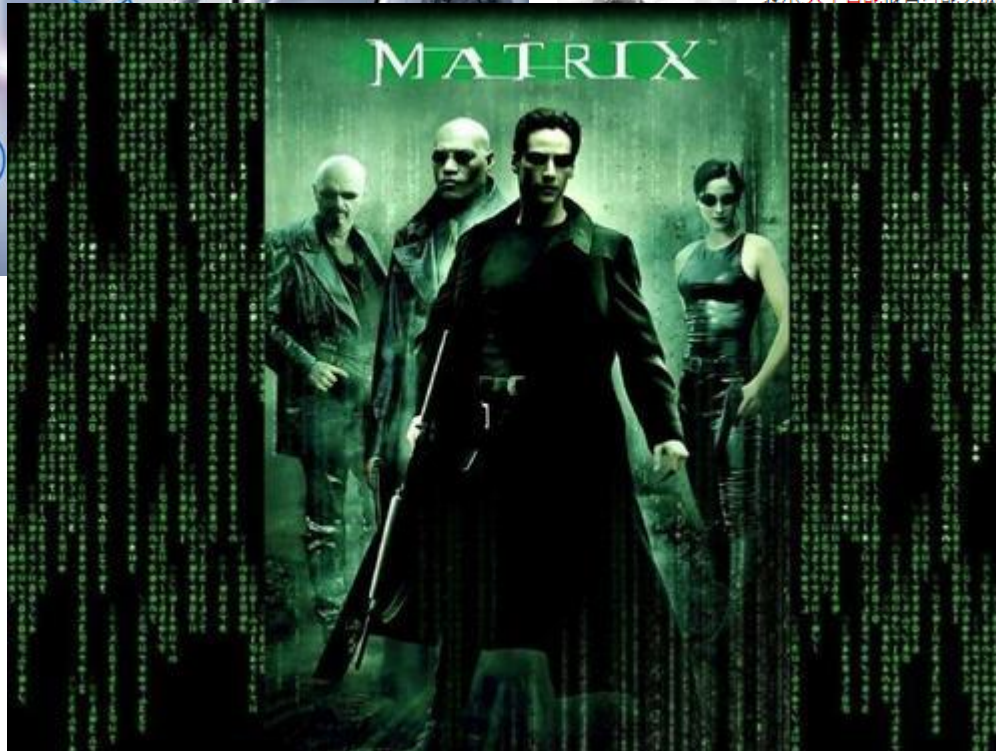


2018年10月8日 - 霍金先生生前曾经对人工智能表示过担忧,霍金先生表示,人工智能很有可能会成为一种不受人类控制的新物种,他们的时... 拥有了自己思维能力,动手...

是否会超越人类?

人工智能是否将会超越人类呢? 人工智能产... 何提高人工智能产品的智能程度是很多研... 见,大多数的研发者都想要...

机大战"谷歌围棋人工智能AlphaGo击败... 能技术破解并超越吗?这一....



理想与现实



选好产品

资产诊断

评测基金

分析涨跌

花呗有利息吗

如果花呗未按时还款，产生逾期，逾期费用=逾期金额*逾期天数*0.05%（建议按期还款，以免产生不良的记录影响信用）。

花呗有利息吗

为你筛选了以下内容：

大v观点 网贷利息那么高，为何还有人借？

我想.....没有人想欠别人钱的，更不想被债务压得喘不过气来！大部分的网贷平台利息都要高于银行，相对也还算比较透明化，但也有不少存在“暗息”的...

06-05 10:29

花呗有利息吗

这个词我还在学，下次告诉你好不好？

咨询其他问题

这个问题好深奥，我要去学习下，你可以问问其他问题~

[illegible]

来源: 澎湃新闻 - 百度快照

科技湃V - 百度快照

<https://baijiahao.baidu.com/s?...> - 百度快照



理想与现实

百度为您找到相关结果约18,900,000个

搜索工具

诺奖得者为何说“人工智能其实就是统计学”AI真的只是华丽辞藻吗



2018年8月12日 - Sargent(托马斯·萨金特)在“共享全球智慧 引领未来科技”世界科技创新论坛上表示:人工智能其实就是统计学,只不过用了个很华丽的辞藻。托马斯·萨金特是谁呢? ...

思维解构 - 百度快照

诺奖得主:人工智能绝不只是统计学!



2018年8月20日 - Sargent在2018年世界科技创新论坛上表示,所有的人工智能利用的都是统计学来解决问题,人工智能只不过是传统统计学挂上了一个很华丽的辞藻,果真如此为什么直到最近56年...

赋能商学院 - 百度快照

任正非说:人工智能就是统计学,你听懂了吗?



2019年2月11日 - 最近任正非在一次采访中说:大学里没有人工智能这门课,人工智能就是统计学,人们老说人工智能、人工智能,而中国的教育却不重视基础教育,尤其不重视数学教育。我个人...

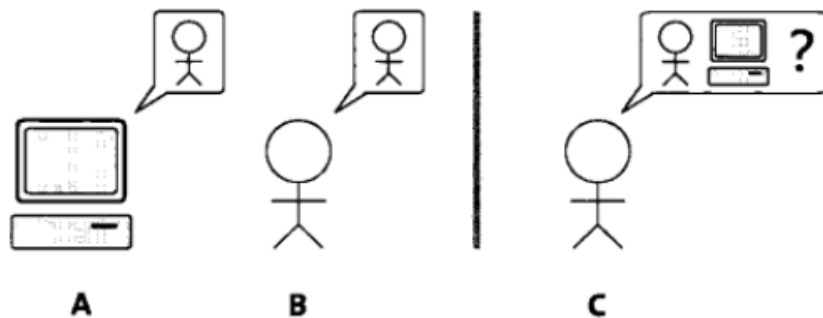
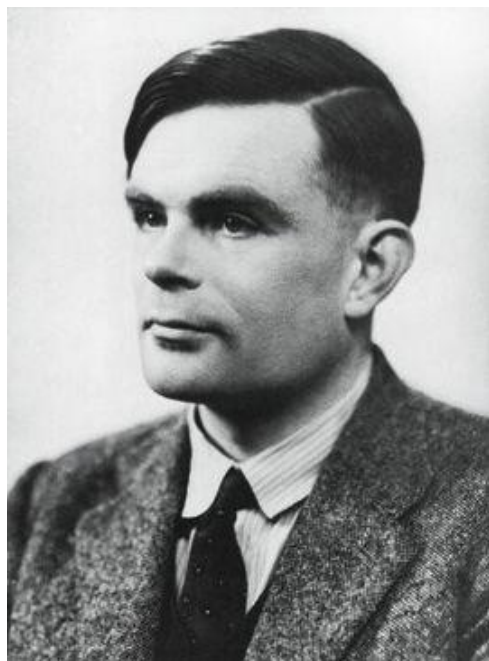
瀑布先生 - 百度快照

任正非:人工智能就是统计学 要提高对统计学重视程度

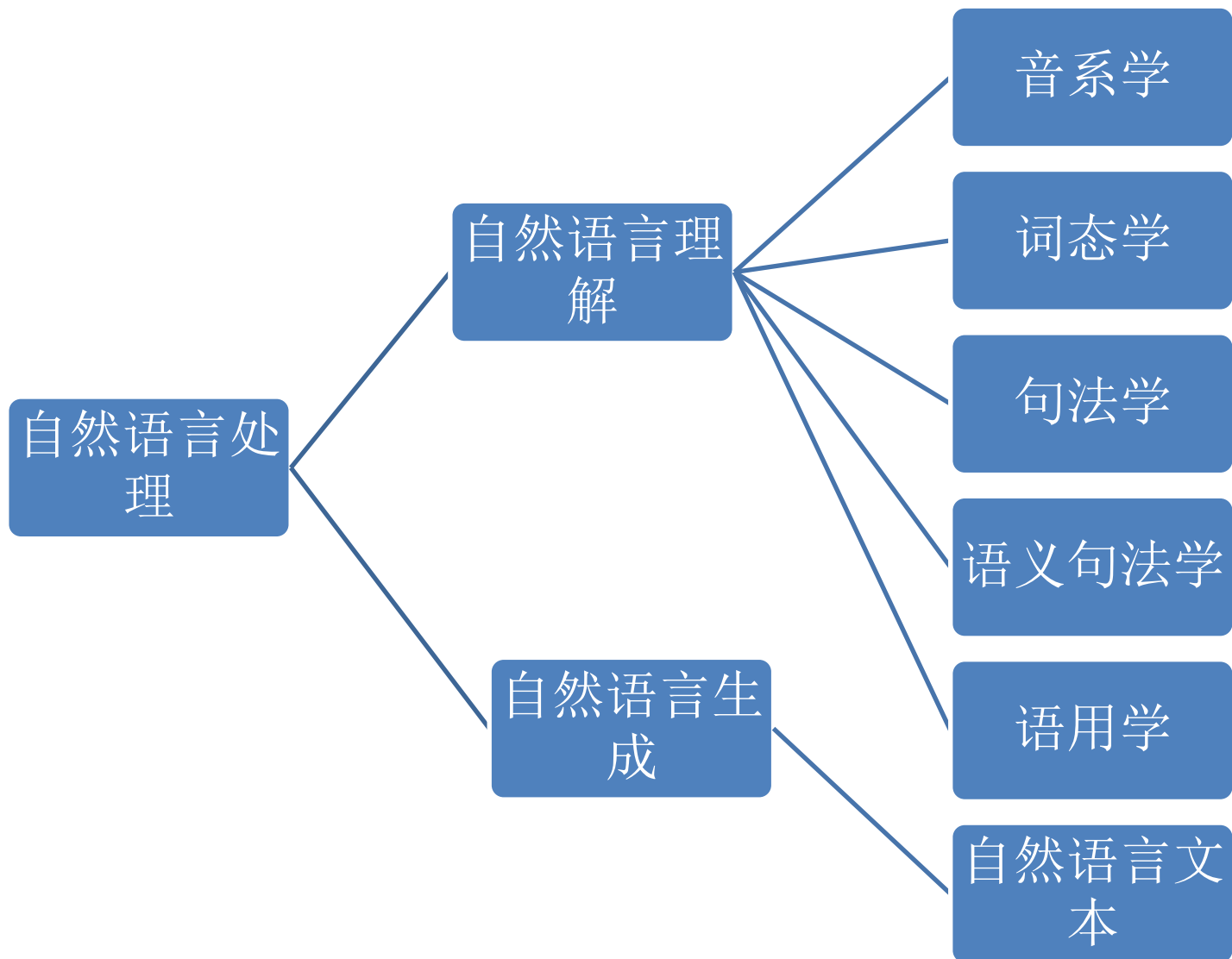
2019年1月20日 - 华为创始人兼CEO任正非在接受央视《面对面》采访,当谈到人工智能是,任正非表示,人工智能就是统计学,这个学科计算机与统计学就是人工智能。任正非指出,...

人工智能的初心

1950年，图灵提出著名的“图灵测试”，指出如果第三者无法辨别人类与人工智能机器反应的差别，则可以论断该机器具备人工智能。



自然语言处理



自然语言处理的研究方向

- 语音识别
- 情感分析
- 文摘生成
- 文本分类
- 知识图谱
- 机器翻译
- 智能问答
- 舆论分析
- 信息检索
- 信息抽取

自然语言处理发展现状

- 《2018年世界人工智能产业发展蓝皮书》预计2021年全球自然语言处理市场的价值可以达到160亿美元。



自然语言处理



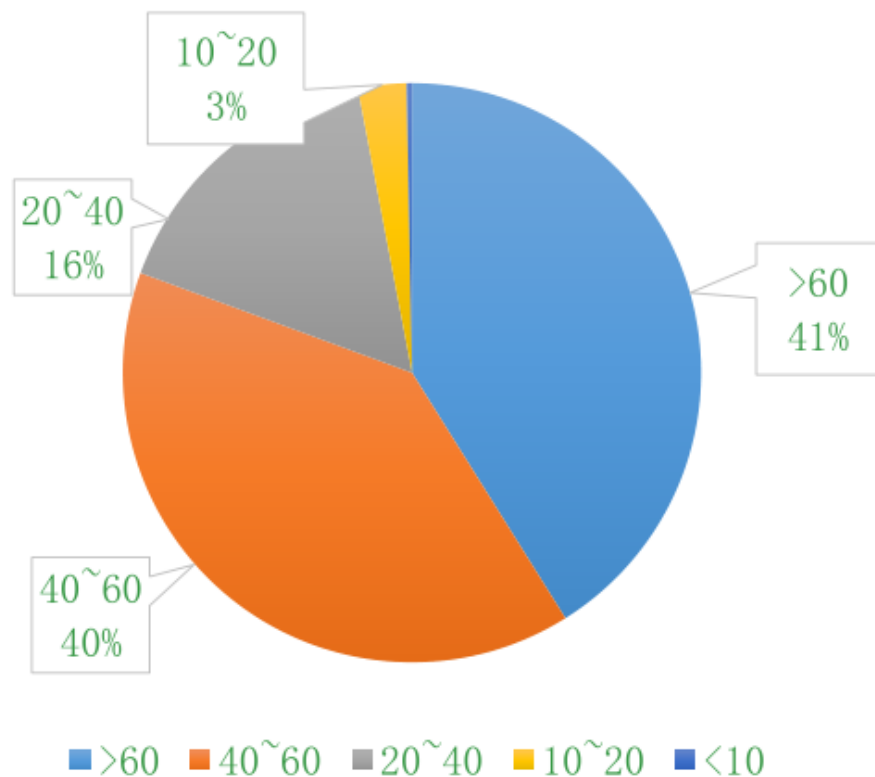
自然语言处理学者全球分布图



自然语言处理华人库专家国内分布



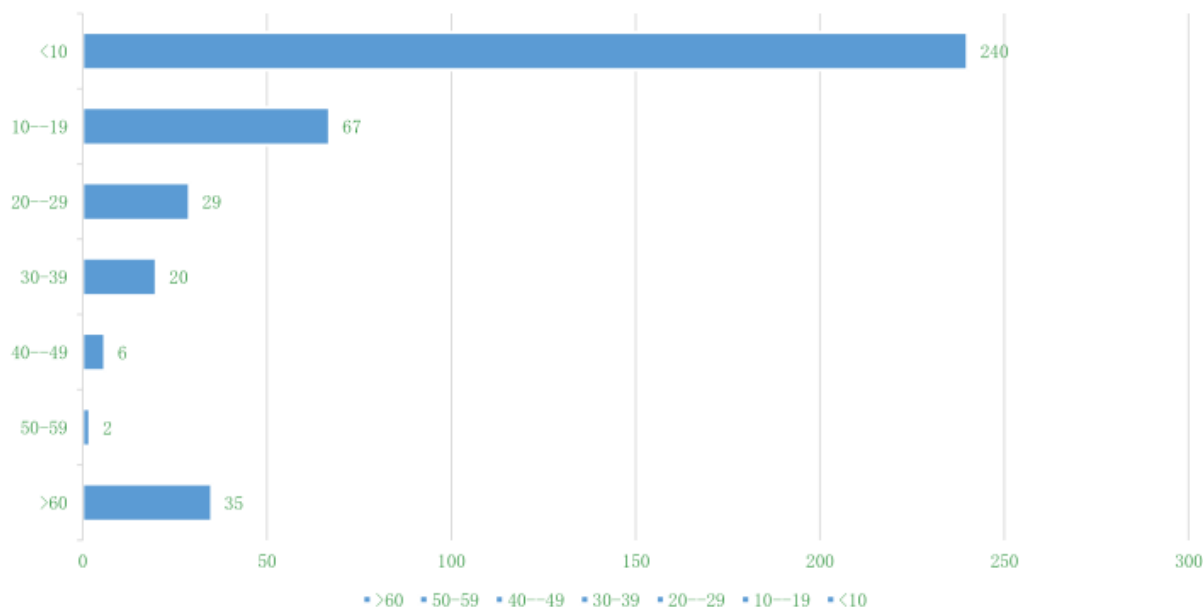
全球专家情况



自然语言处理顶尖学者h-index分布

- 全球自然语言处理顶尖学者的h-index平均数为**59**
- h-index指数大于60的学者最多占比**41%**
- h-index指数在40到60之间的学者次之，占比**40%**

中国专家情况

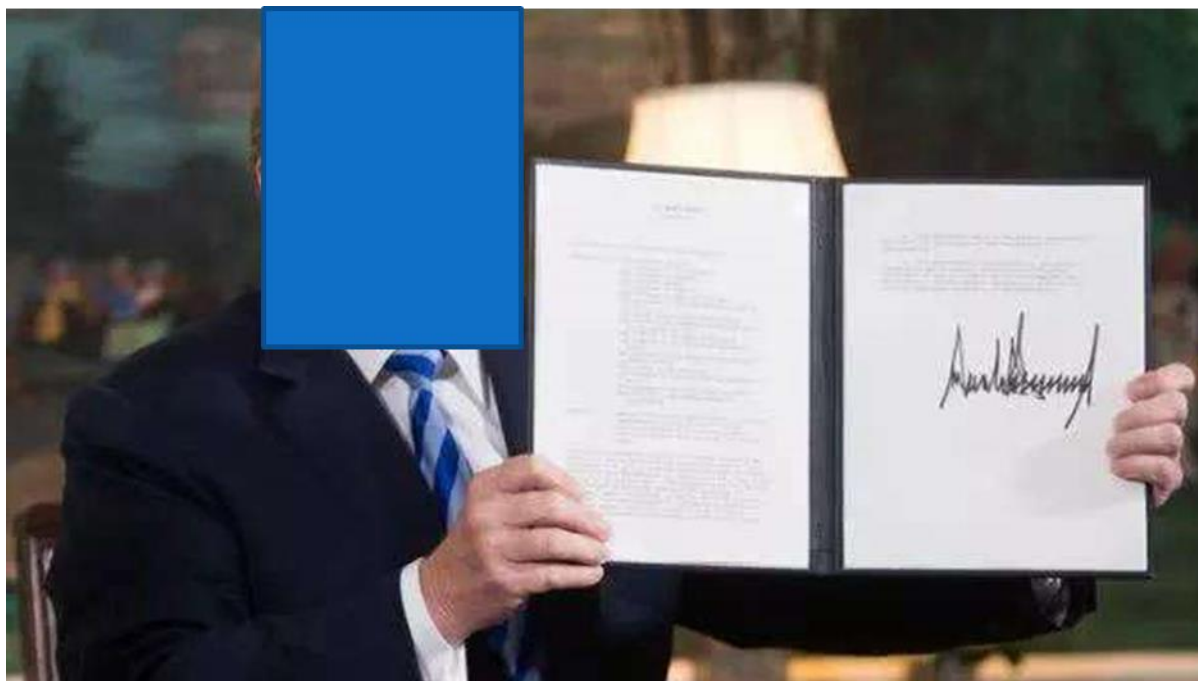


自然语言处理华人库专家h-index分布

- h-index指数的平均数为14, 这一数值远远低于自然语言处理全球top1000学者h-index指数平均数。
- h-index指数<10的专家人数最多, 占比60%
- 10-19次之, 占比17%
- >60的专家占比仅占9%
- 自然语言处理华人专家整体水平低于自然语言处理领域全球top1000的学者, 尤其是在h-index指数>60的学者方面有所欠缺。

我们要研究的

- 对于英语来说，词与词之间有明确的分界符(手写体除外)，但汉语的词与词之间却没有明确的界限，所以需要有一个方法对中文进行分词。



中文分词

结婚的和尚未结婚的

结婚的和尚未结婚的



网页 资讯 视频 图片 知道 文库 贴吧 采购 地图 更多»

百度为您找到相关结果约4,010,000个

搜索工具

[佛祖没说过和尚不能结婚,和尚的老婆曾有专门称呼_搜狐](#)



2018年8月4日 - 在我们的认知里,和尚是不能结婚的。但史料记载,后秦高僧鸠摩罗什一生就结过两次婚,并且还生了孩子。据说第二次结婚后,在佛教界产生了很大的躁动。...

搜狐网 - 百度快照

[为什么和尚不许结婚,可释迦牟尼佛却结婚生子](#)



2018年9月28日 - 我们都知道,佛教有规定,和尚、尼姑都不能结婚。可在佛教史上,却有一个让和尚尼姑们尴尬却又无法回避的事实,那就是释迦牟尼佛结了婚,生了子。这件事情被佛徒...

潭映大千 - 百度快照

[和尚不能结婚道士能结婚为什么 这其实有一定的误解_秀目网](#)



2018年1月11日 - 关于和尚为什么不能结婚,其实有过争辩。有人曾经提过要不要学下日本僧人可以结婚,不过最后这个提案被否决了。而关于道士为什么不能够结婚,其实也是一个...

秀目网 - 百度快照

[法律有规定和尚不能结婚吗?_百度知道](#)

4个回答 - 回答时间: 2017年9月5日

最佳答案: 当然允许。我国婚姻法第五条 结婚必须男女双方完全自愿,不许任何一方对他方加以

隐马尔可夫模型

- 一个人的身体有正常、轻感冒、重感冒三种状态，某个小镇的医疗水平，无法测出某个人到底是正常，还是轻感冒，还是重感冒。他们只能看出来这个人到底是没什么症状、咳嗽、发烧还是拉肚子。假设病人的外在表现仅仅和他当天的内在状态有关，和其他的条件无关，关系如下表：

	活蹦乱跳	咳嗽	发烧	拉肚子
正常	0.7	0.2	0	0.1
轻感冒	0.5	0.2	0.2	0.1
重感冒	0.3	0.2	0.3	0.2

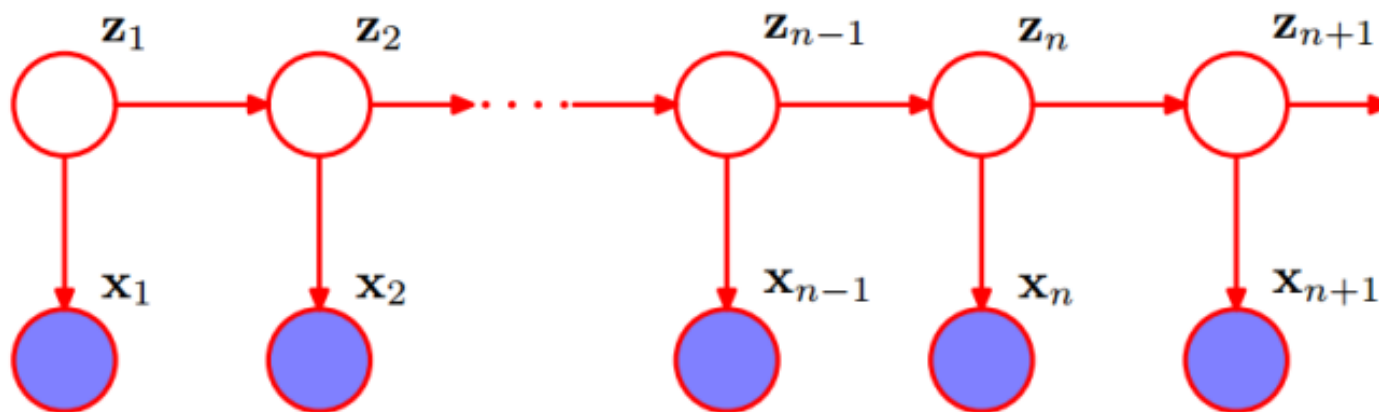
隐马尔可夫模型

- 同时，这个小镇有53%的人身体是正常的，31%的人患了轻感冒，而16%的人患了重感冒，那么一个人如果我们不知道任何信息的情况下去猜，显然我们应该猜这个人53%的可能性是正常的，31%的可能性患了轻感冒，而16%的可能患了重感冒。同时一个人两天之间的身体状况满足下面的关系：

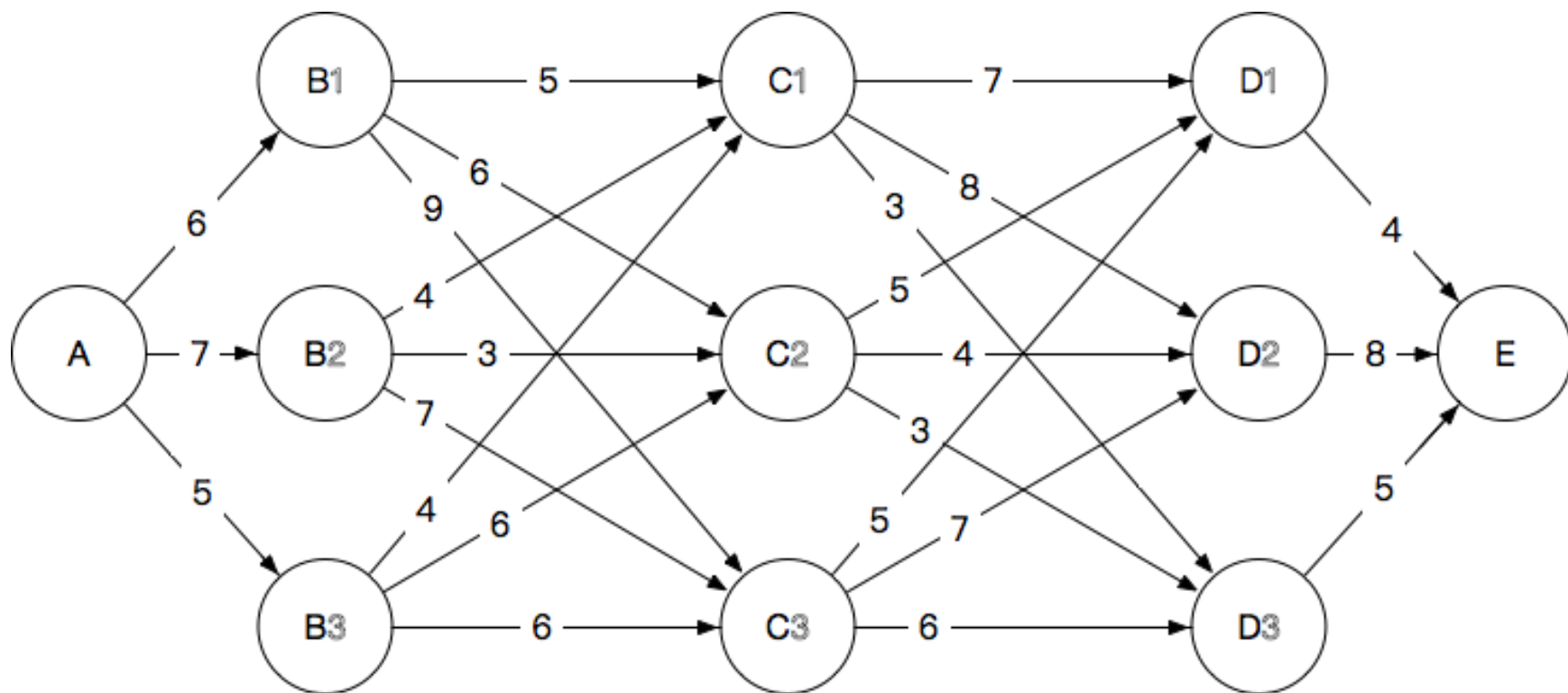
		今天		
昨天	状态	正常	轻感冒	重感冒
	正常	0.7	0.2	0.1
	轻感冒	0.4	0.4	0.2
	重感冒	0.2	0.5	0.3

隐马尔可夫模型

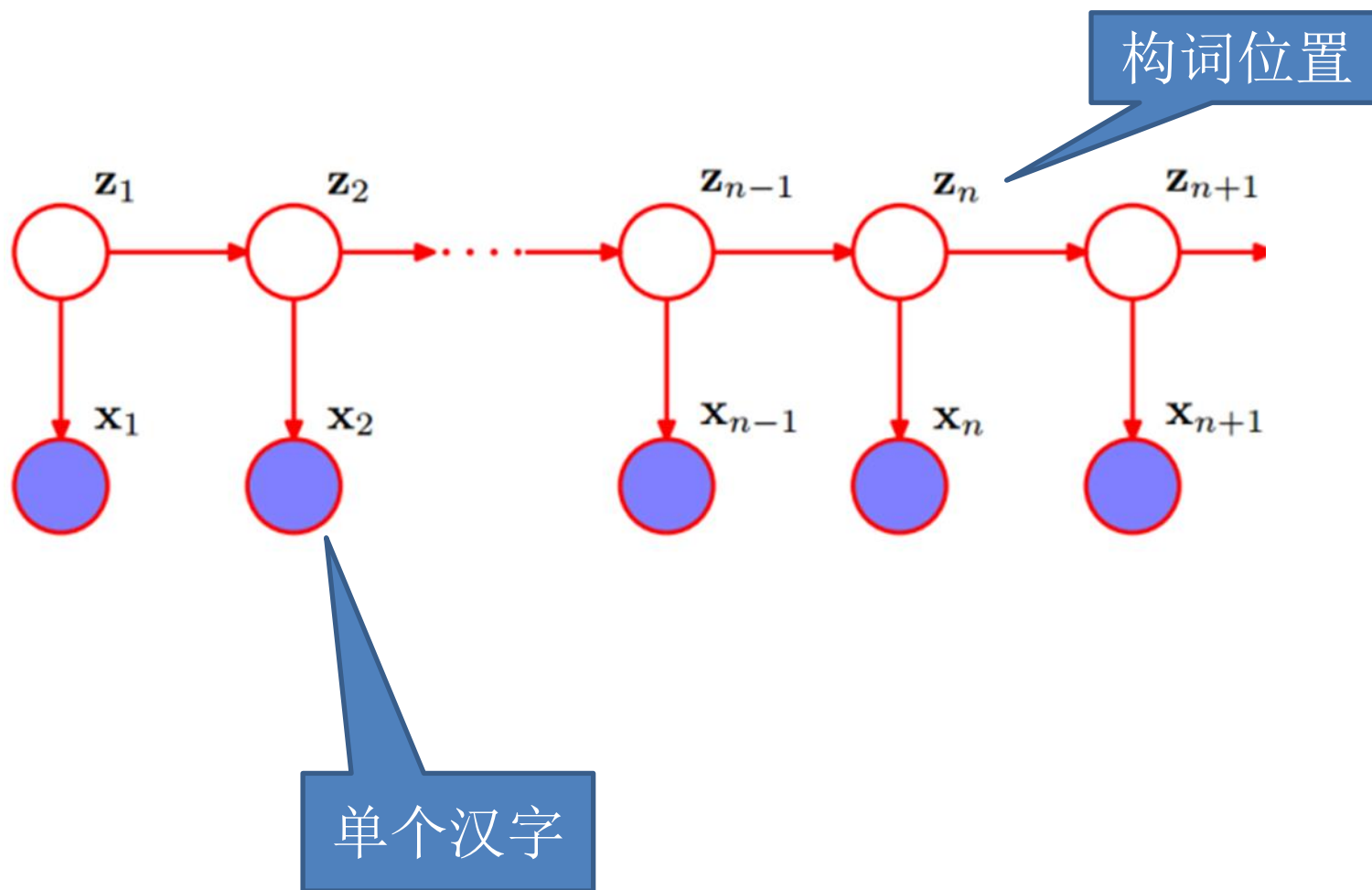
- z 表示身体状况(正常还是感冒)， x 表示外在表现(咳嗽等)， z 是无法直接观测的，但是 z 满足马尔可夫链的性质。



维特比算法



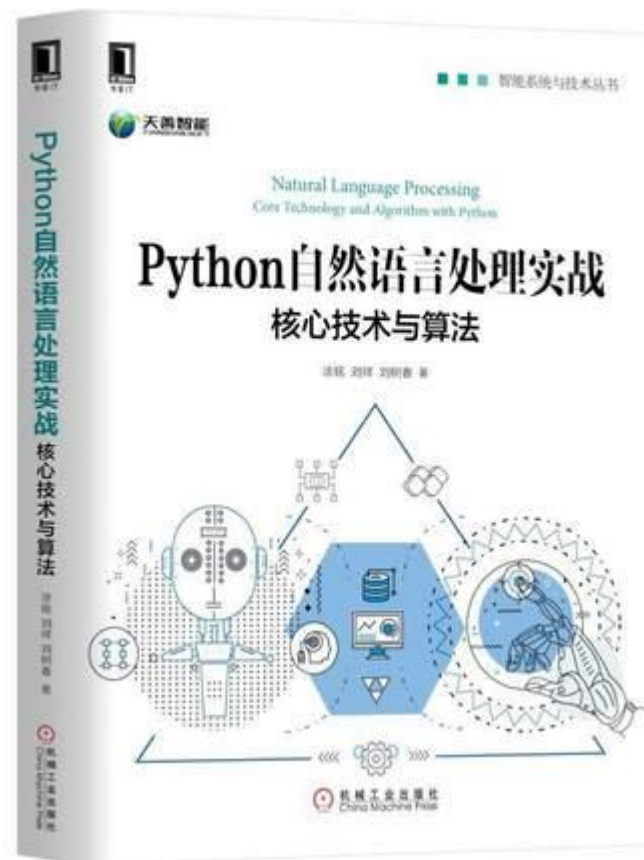
隐马尔可夫模型与中文分词



总结

- 数学有益于身体健康
- 逻辑回归和神经网络基础
- 图聚类：谱聚类和马尔可夫聚类
- AlphaGo Zero：蒙特卡罗方法
- 自然语言处理：隐马尔可夫模型

参考资料(排名不分先后)



参考资料(排名不分先后)

- 知乎
- CSDN
- GitHub
- 鸠摩搜书
- 小不点搜索<http://www.xiaoso.net/>
- <http://www.allitebooks.org/>