

认知领域的剩余价值



高 扬
2016/9

欢聚时代



目录



- 1、剩余价值的体现
- 2、大数据框架的作用
- 3、机器学习算法
- 4、深度学习趣用

剩余价值的体现



- 剩余价值

- 剩余价值=（出售价格—成本）

- 出售价格：受到市场供需平衡影响
 - 成本：用机器代替人



剩余价值的体现



- 大数据的剩余价值
 - 通过消除不确定性减少试错成本



剩余价值的体现



- 统计——认知的最朴素方式
 - SUM,AVG,COUNT——SQL中的统计

剩余价值的体现



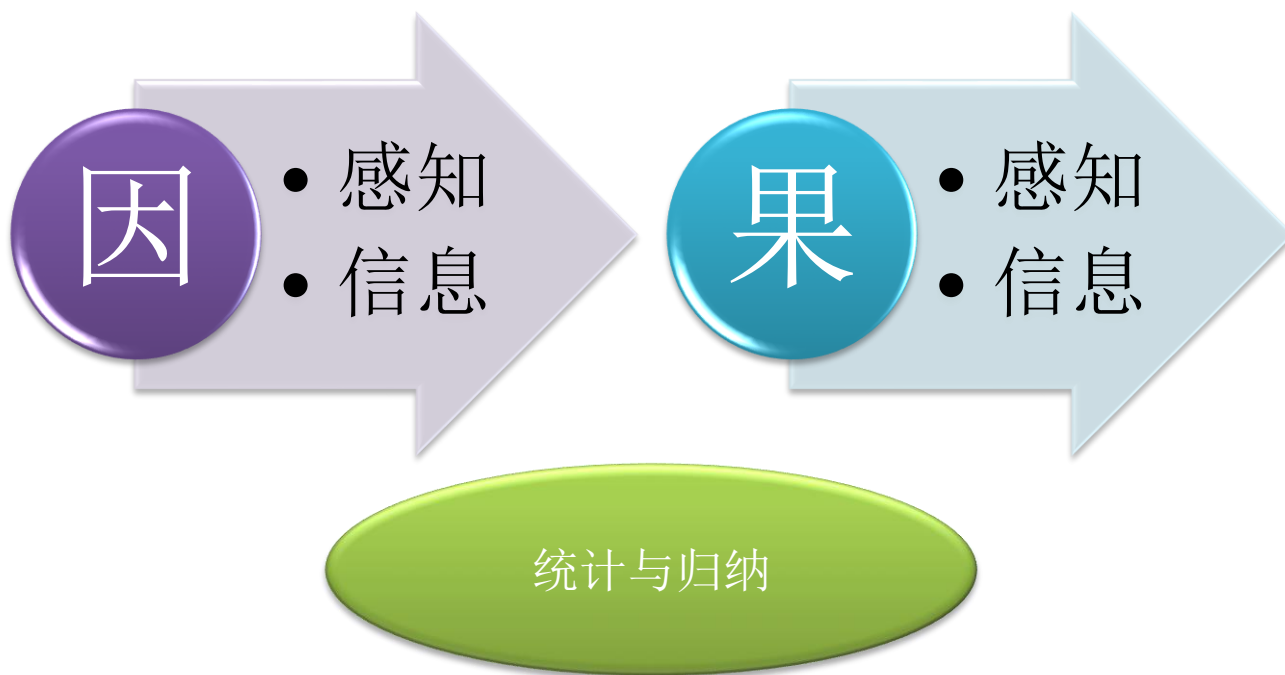
- 统计——认知的最朴素方式
 - 现实生活中的统计，因与果（关联）
 - 天上下雨地上流
 - 冰能不能燃烧



剩余价值的体现



- 统计——认知的最朴素方式
 - 统计是人类在最朴素的条件下所进行的认知行为



大数据框架的作用



- Hadoop
 - HDFS: 解决永久存储的低成本扩容问题
 - MapReduce: 解决分布式计算框架问题
 - 作用: 降低分布式存储与计算的实现成本和落地成本



大数据框架的作用



- Spark
 - Tachyon: 内存里的HDFS
 - RDD: 数据块句柄
 - Spark Streaming: 准实时计算
 - 作用: 降低准实时计算以及内存计算的实现与落地成本

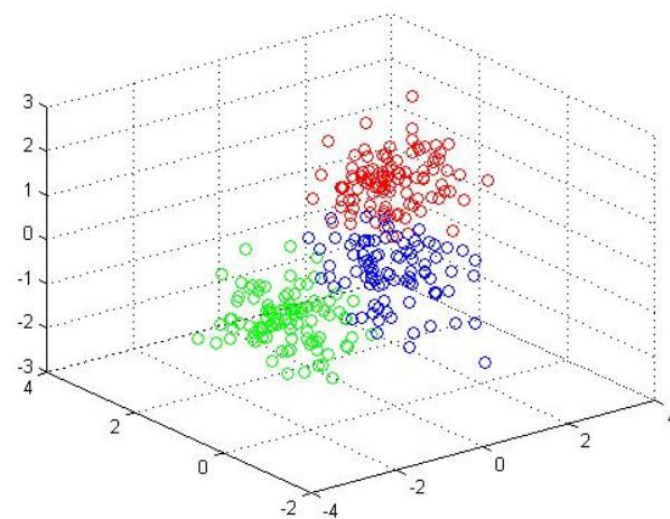
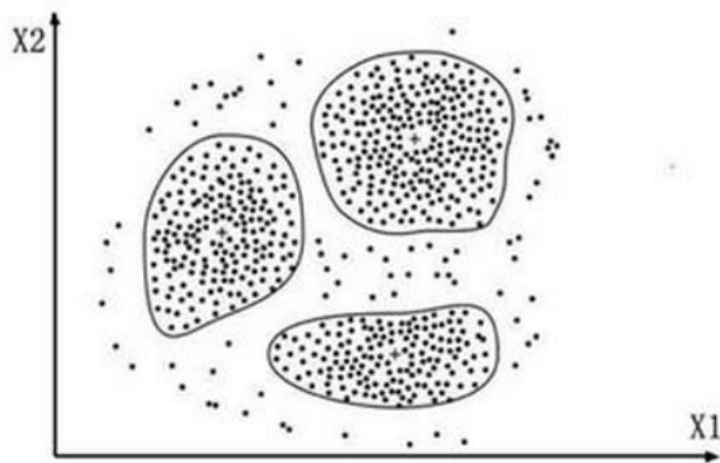


机器学习算法



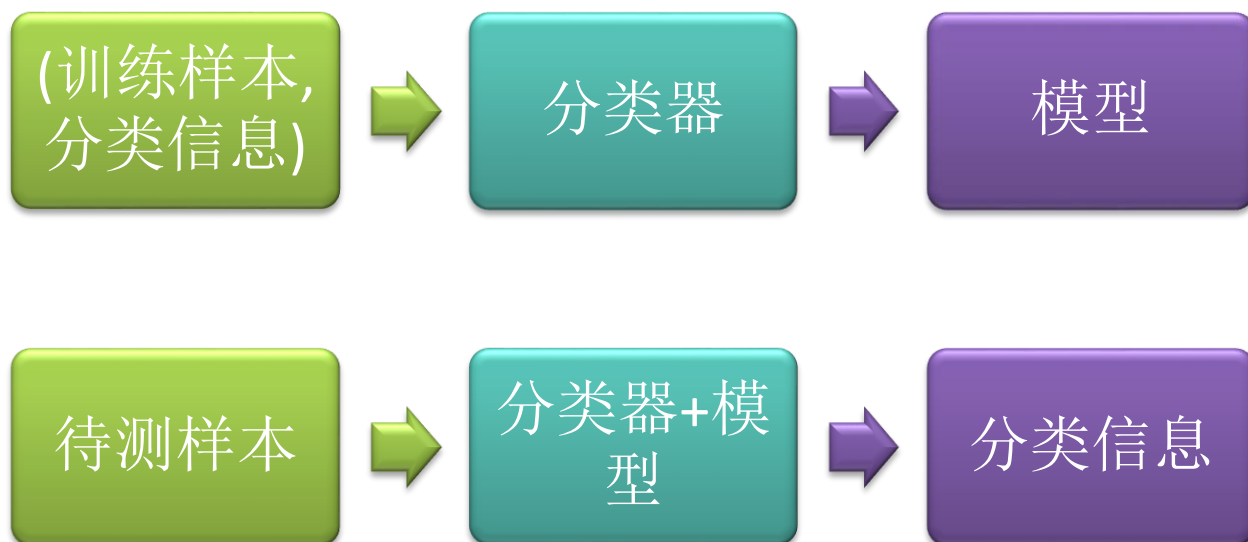
- 核心目的：统计、归纳，消除不确定性
 - 无监督学习（聚类）
 - 有监督学习（分类）

- 无监督学习（聚类）
 - K-Means、K-Means++
 - 聚类与离群点



- 有监督学习（分类）

- Step1: 训练
- Step2: 工作



- 有监督学习（分类）

- 朴素贝叶斯

- 事情发生的概率，会由于知晓一些事情发生的概率而发生变化

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

- 有监督学习（分类）

- 决策树

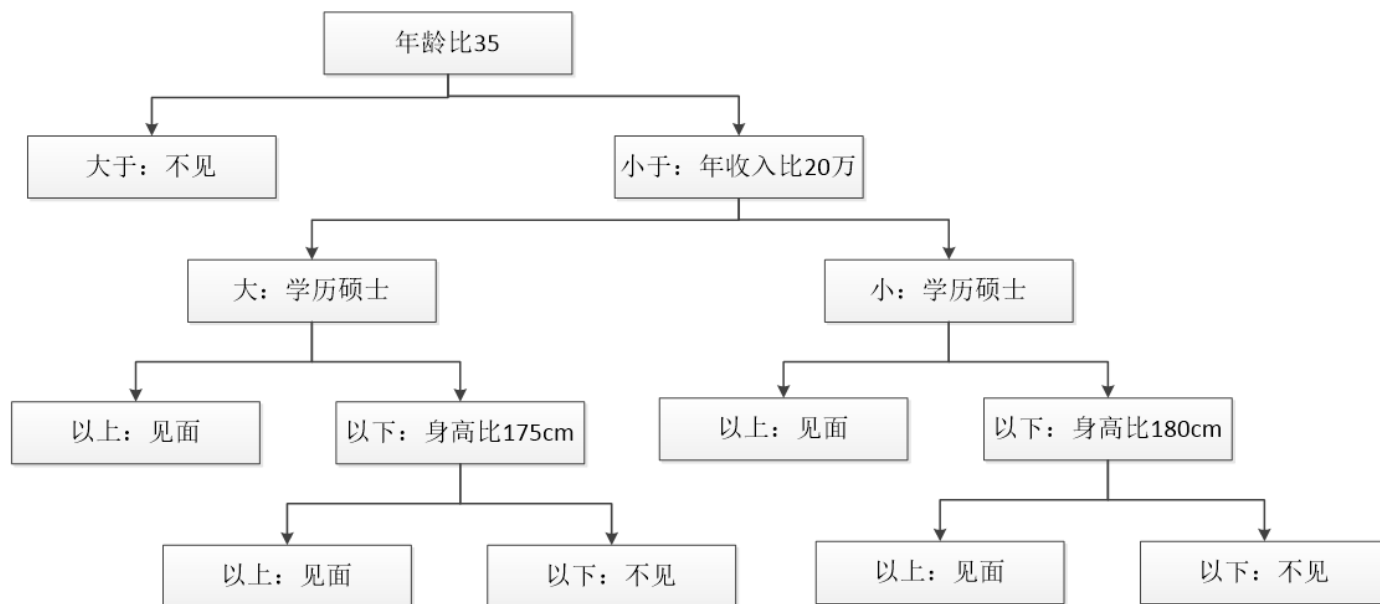
- 通过尽可能少的条件得到尽可能大的分类纯度

网站 ID	年龄（岁）	身高（cm）	年收入（万）	学历	有否相亲
XXXXXXX	25	179	15	大专	N
XXXXXXX	33	190	19	大专	Y
XXXXXXX	28	180	18	硕士	Y
XXXXXXX	25	178	18	硕士	Y
XXXXXXX	46	177	100	硕士	N
XXXXXXX	40	170	70	本科	N
XXXXXXX	34	174	20	硕士	Y
XXXXXXX	36	181	55	本科	N
XXXXXXX	35	170	25	硕士	Y
XXXXXXX	30	180	35	本科	Y
XXXXXXX	28	174	30	本科	N
XXXXXXX	29	176	36	本科	Y

- 有监督学习（分类）

- 决策树

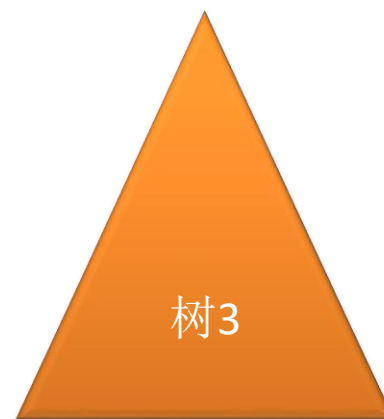
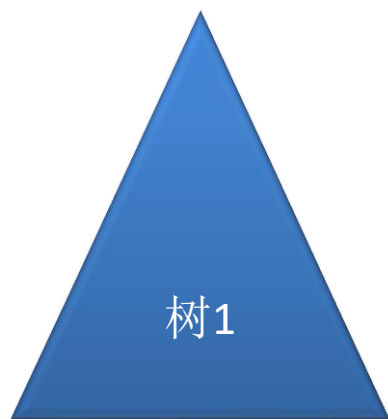
- 通过尽可能少的条件得到尽可能大的分类纯度



- 有监督学习（分类）

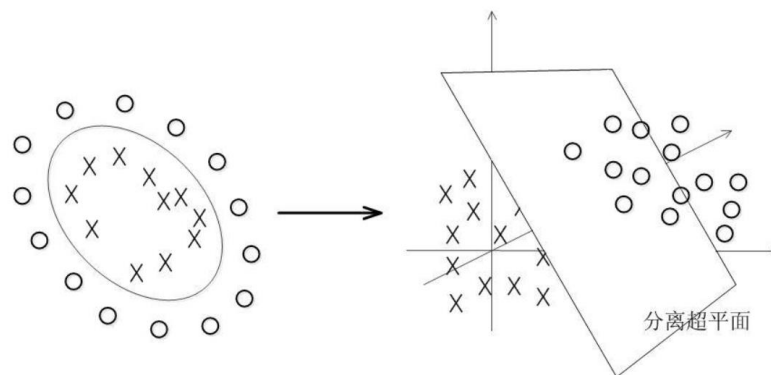
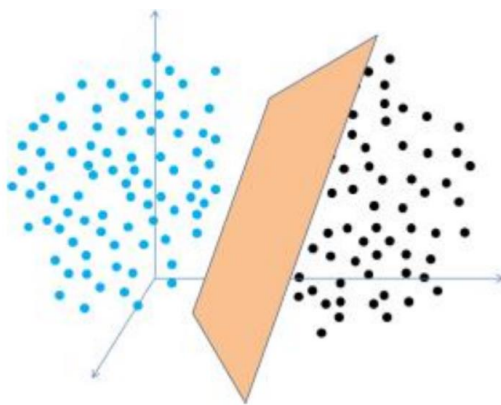
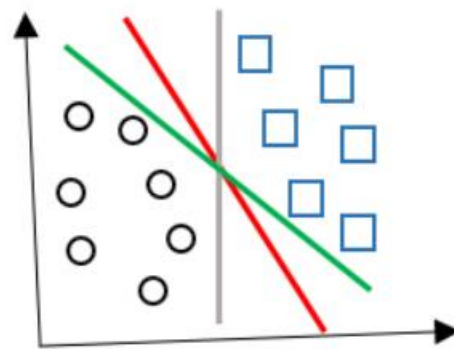
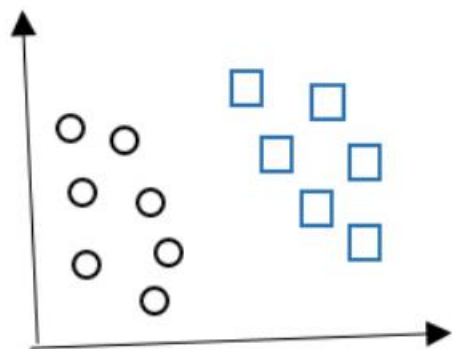
- 随机森林

- 用多个浅层的分类不纯的决策树投票表决，以概率的方式裁决出分类最大可能性



- 有监督学习（分类）
 - 支持向量机SVM
 - 能用线分开的就分开，分不开升个维再分开
 - 核函数：升维利器

- 有监督学习（分类）
 - 支持向量机SVM

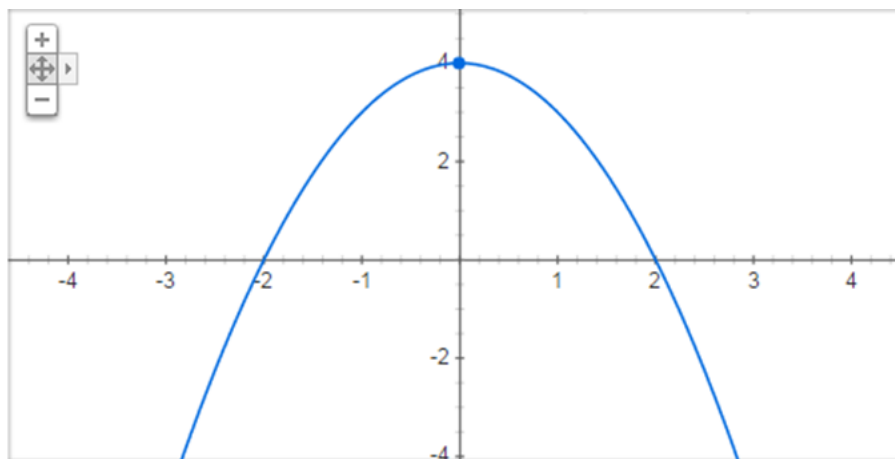


机器学习算法



- 有监督学习（分类）
 - 支持向量机SVM
 - $[-2,2] \rightarrow 1$ ，其余 $\rightarrow 0$

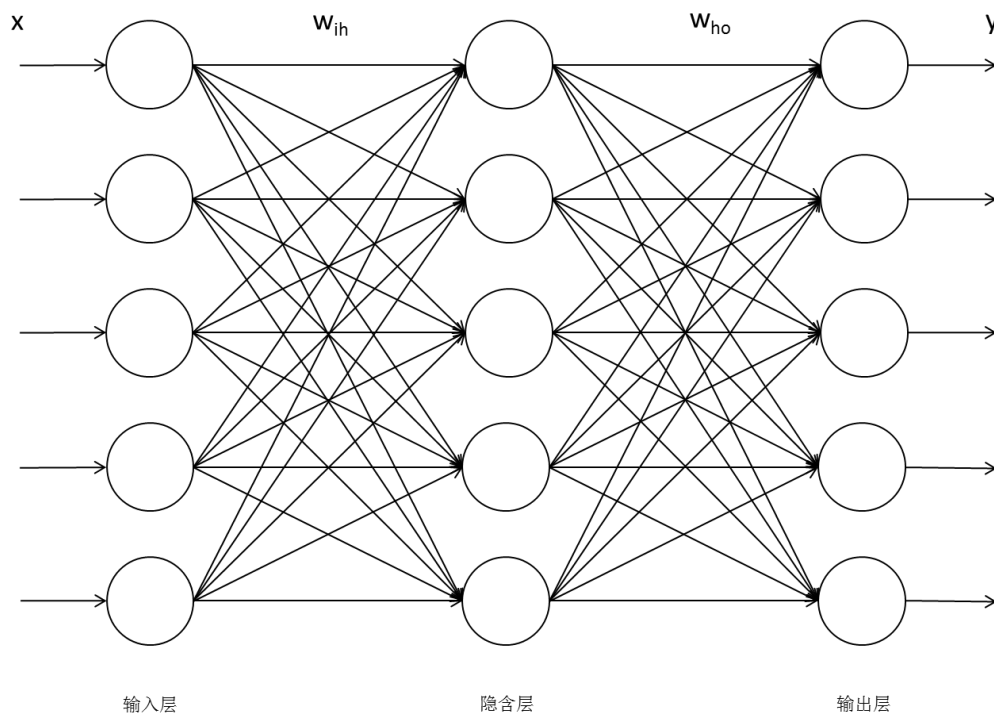
$$f(y) = \begin{cases} 1 & y > 4 \\ 0 & y \leq 4 \end{cases} \quad \begin{matrix} y = -x^2 + 4 \\ z = f(y) \end{matrix}$$



深度学习趣用

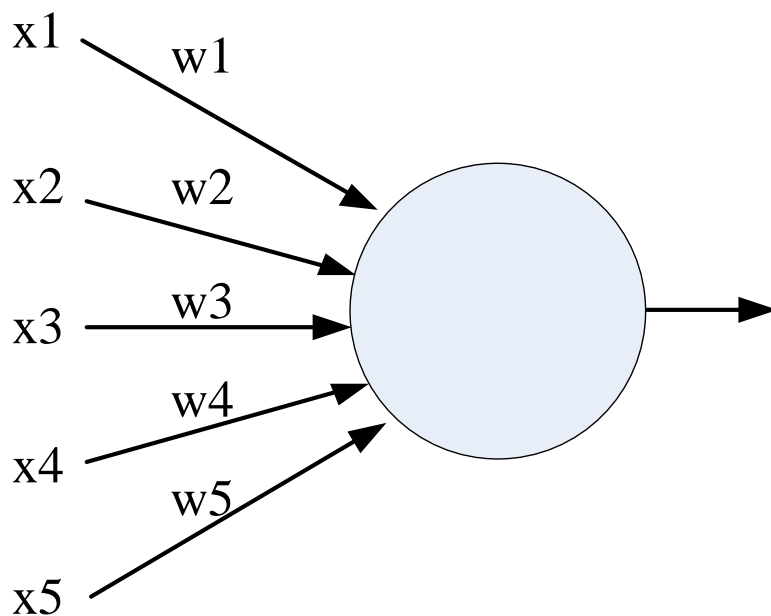


- 神经网络为什么那么强
 - 基本结构
 - $F(x)=wt+b$



- 神经网络为什么那么强

- $F(x) = w_1x_1 + w_2x_2 + w_3x_3 + w_4x_4 + w_5x_5 + b$

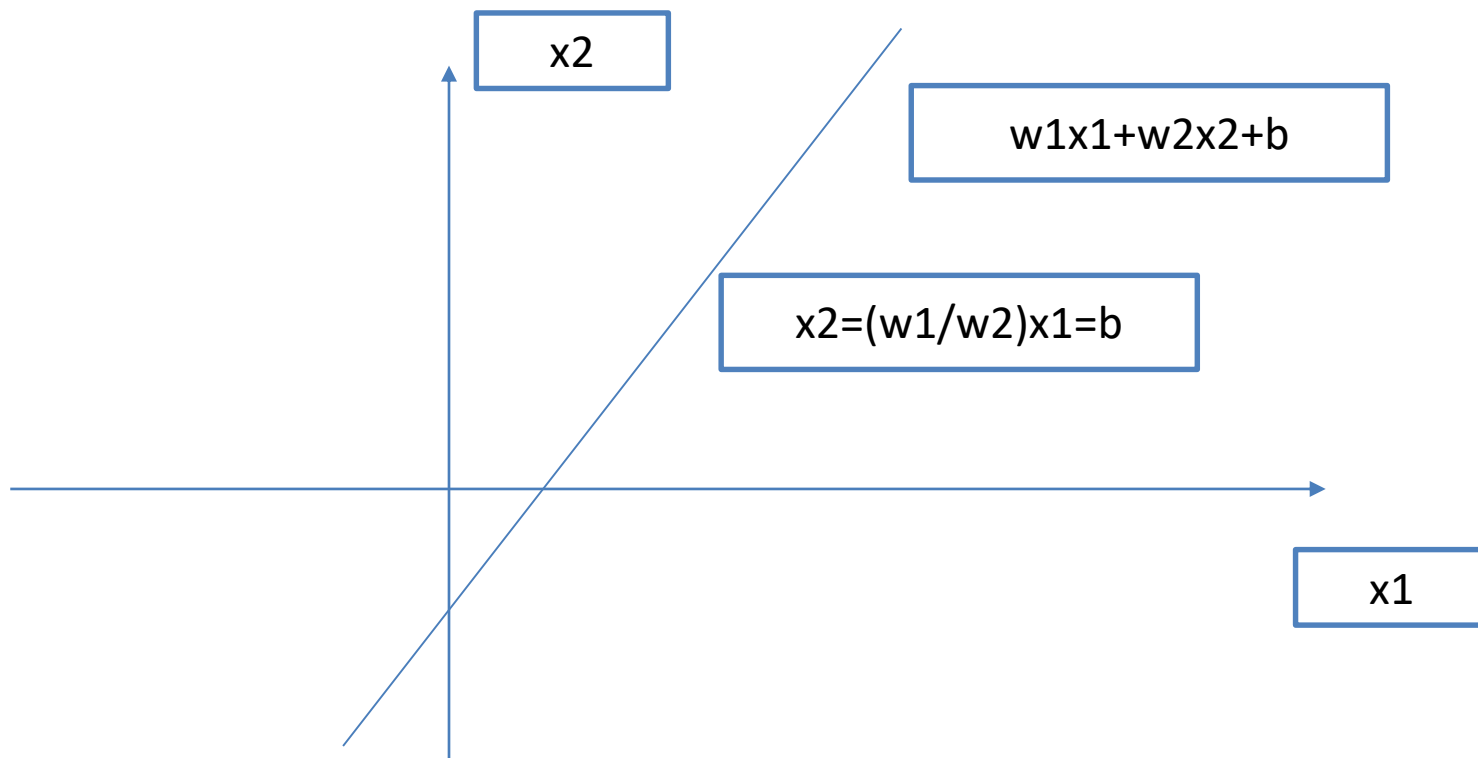


深度学习趣用



- 神经网络为什么那么强

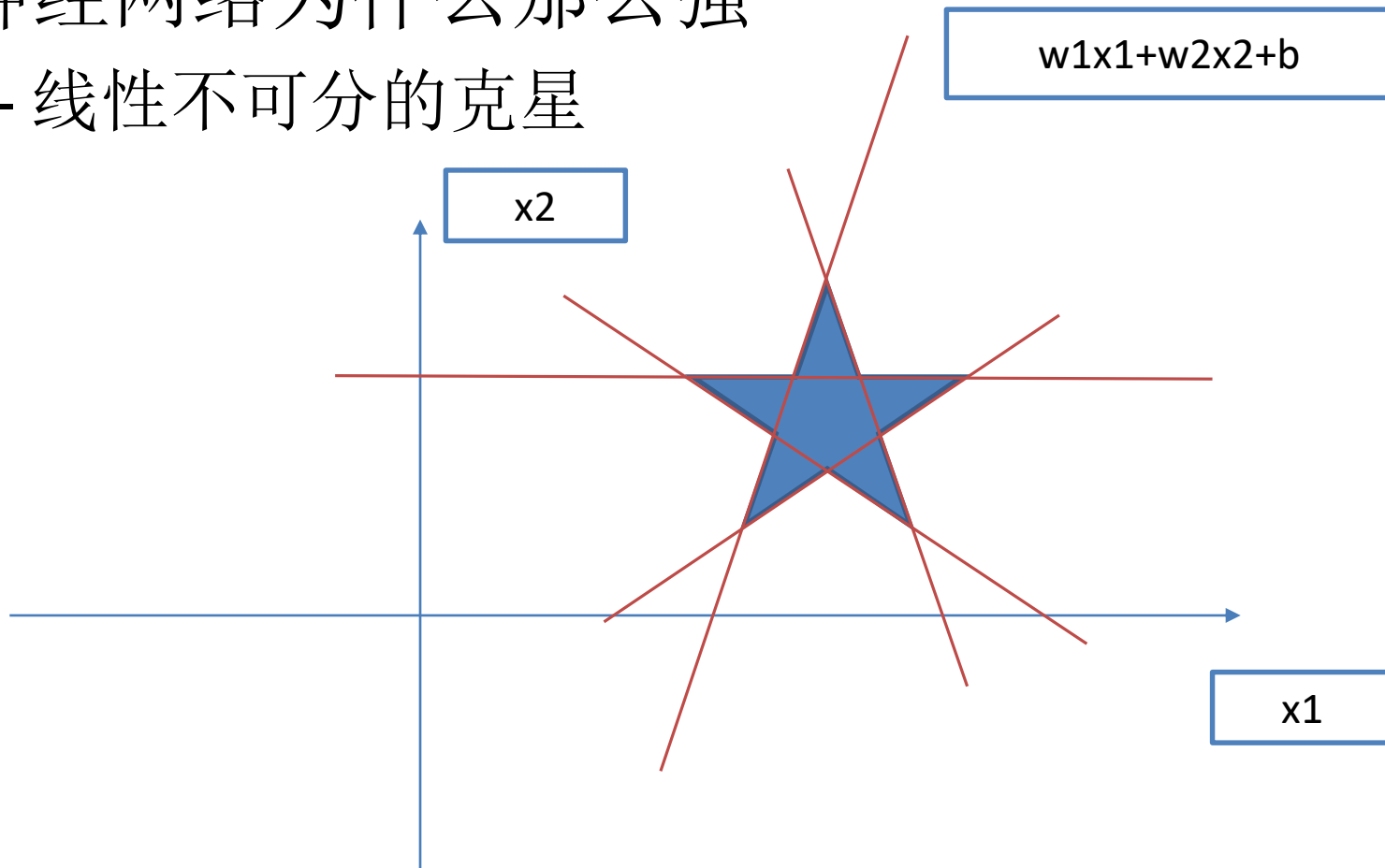
– $F(x) = w_1x_1 + w_2x_2 + b$



深度学习趣用



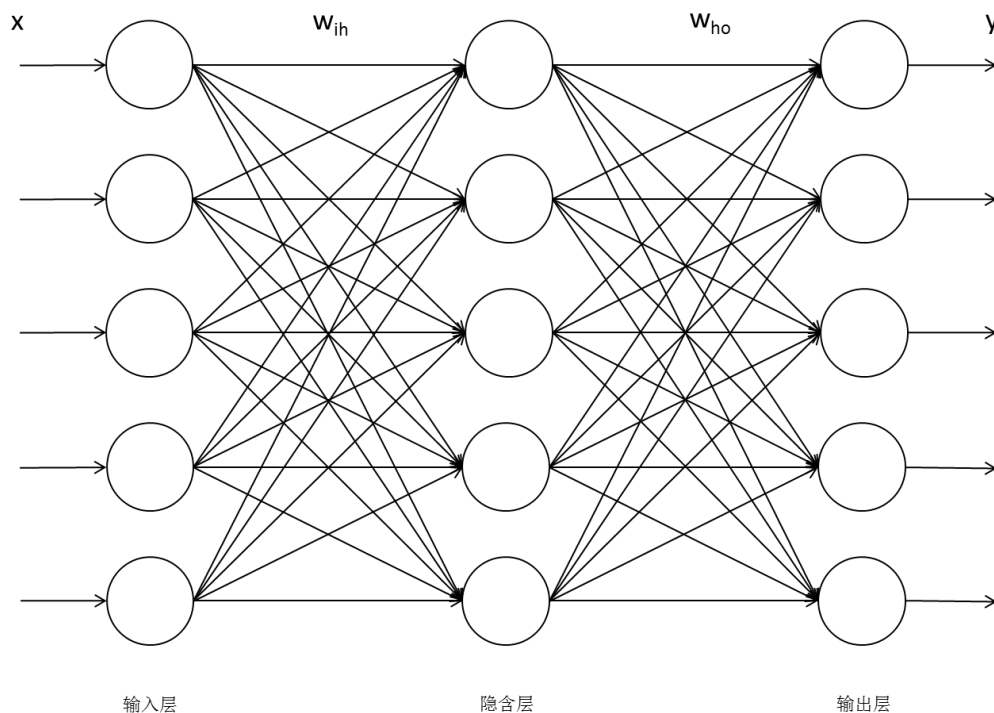
- 神经网络为什么那么强
— 线性不可分的克星



深度学习趣用



- 神经网络为什么那么强
 - 通过训练找到所有的保证分类最为准确的 w



深度学习趣用



• 趣味应用

— RNN — 作诗姬

稻香居电脑作诗机网络版5. xxN >>>>> 电脑作诗 [\[发表评论/查看评论\]](#)

请选择韵部（本测试版限制使用十个韵部，完整版需下载本地硬盘版）：

☐ 一东 ☐ 四支 ☐ 五微 ☐ 十灰 ☐ 十一真
☐ 一先 ☐ 七阳 ☐ 八庚 ☒ 十一尤 ☐ 十二侵

请选择体裁：☐ 五绝 ☐ 五律 ☐ 七绝 ☒ 七律

请输入命题： （也可不输入命题，电脑作《无题》）

命题不成功，电脑自由作诗。

《无题》

槐花塞北望神州，暮雨关山猿独愁。
称近晴光随意落，宜裁佳气任情流。
听猿过寺吟诗入，避宴看山折笋游。
绕树闲吟休问瘦，梅黄子细野池幽。

结果：存在4%的词汇无法生成正确的关联。您可以返回再试几次。

自由作诗结果：

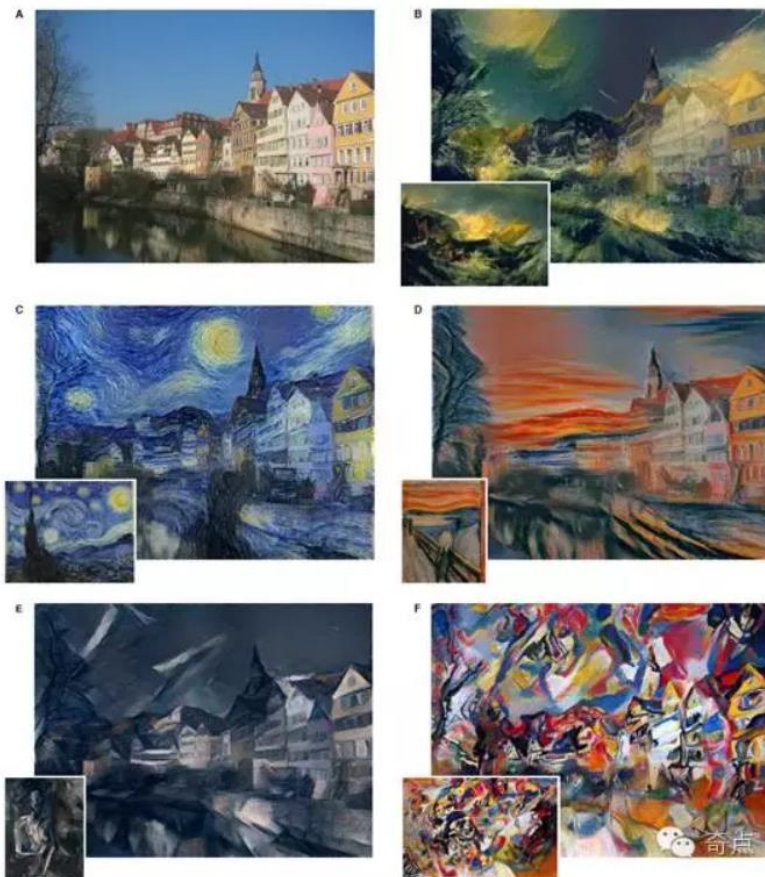
命题成功。

《美景》

路上他乡便去游，花开且伴钓中流。
犹然春榜尝新酒，几逐一家上小舟。
相倚飞舆伤岁暮，可能行殿乱乡愁。
江东领事无因见，梦觉琴堂暮未休。

结果：存在29%的词汇无法生成正确的关联。您可以返回再试几次。

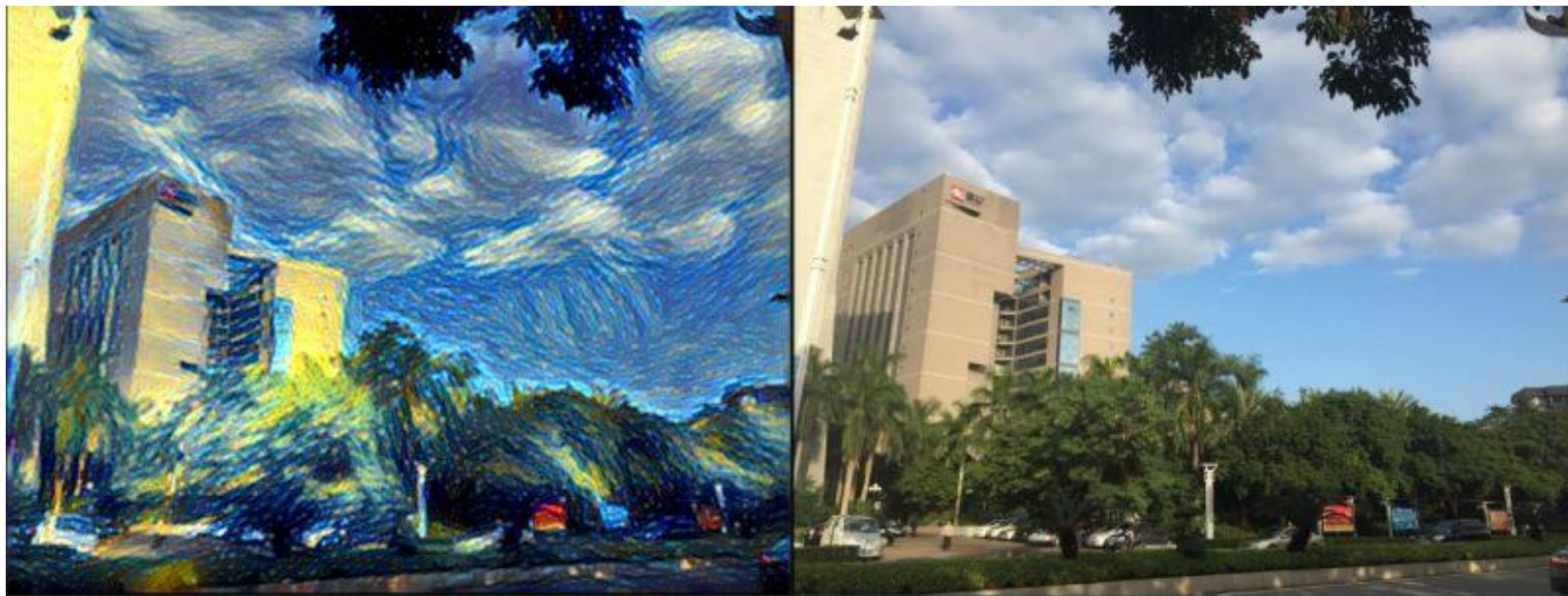
- 趣味应用
 - CNN — 绘画风格临摹



深度学习趣用



- 趣味应用
 - CNN — 绘画风格临摹



小广告时间



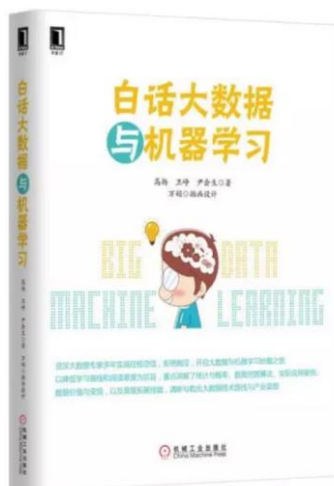
- 新书推荐

- 《白话大数据与机器学习》

- 已由机械工业出版社出版

- 《白话深度学习与Tensorflow》

- 预计明年上半年由机械工业出版社出版



The end,
thanks!

