

人智已

<YL2018.6>

考点汇总

1 α - β 剪枝 / 极大极小 ✓

2 A^* 搜索 ✓

3 SVM (线性/非线性) ✓

4 决策树 ✓

5 归结 / 逆向演绎 ✓

6 Bayes

7 神经元 ✓

8 退火, 遗传

机器学习 · 第四章

朴素贝叶斯 Naive Bayes

支持向量机 SVM

决策树

朴素贝叶斯 Bayes

输入: $X_i \in \mathbb{R}^n$, $Y_i \in \{C_1, C_2, \dots, C_K\}$
特征向量 分类

公式:

$$P(Y=C_k | X=x) = \frac{P(X=x | Y=C_k) P(Y=C_k)}{\sum_i P(X=x | Y=C_i) P(Y=C_i)}$$

独立性:

$$P(X=x | Y=C_k) = \prod_{i=1}^n P(X^{(i)} = x^{(i)} | Y=C_k)$$

朴素贝叶斯 Bayes

分类器 由于分母一样, 故令分子最大即可

$$y = \underset{C_k}{\operatorname{argmax}} P(Y=C_k) \prod_{j=1}^n P(X^{(j)}=x^{(j)} | Y=C_k)$$

参数 $P(Y=C_k)$ = 样本中 C_k 的比例

$P(X^j=x^j | Y=C_k)$ = 样本中 C_k 且满足 $X^j=x^j$ 的比例

朴素贝叶斯 Bayes

拉普拉斯平滑 $\lambda=1$

$$P(X^j = x^j | Y = c_k) = \frac{\sum I(x_i^j = x^j, y_i = c_k) + \lambda}{\sum I(y_i = c_k) + S_j \lambda}$$

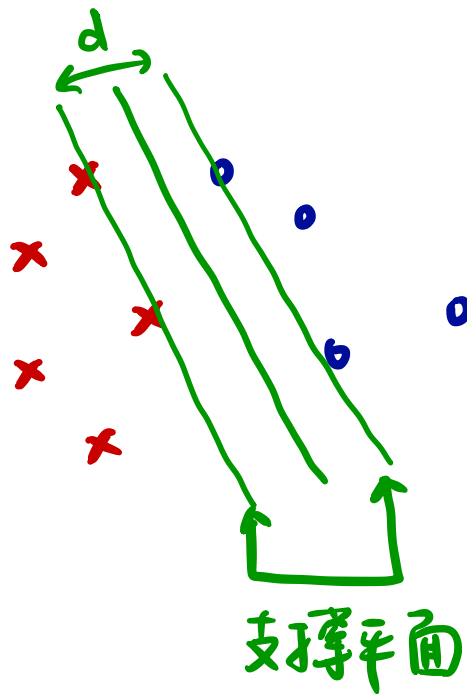
$$P(Y = c_k) = \frac{\sum I(y_i = c_k) + \lambda}{N + K \lambda}$$

例子

看PPT.

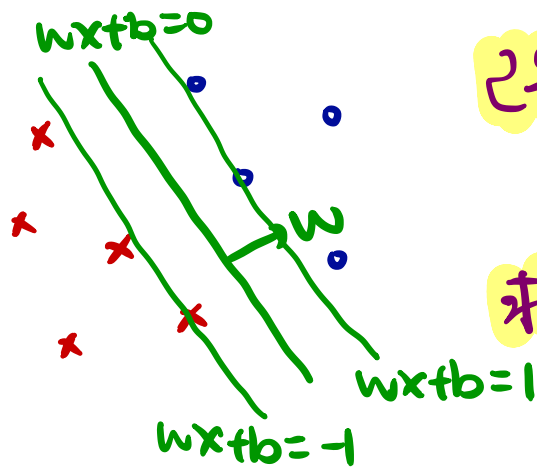
SVM支持向量机

三类问题: 线性 / 线性不可分 / 非线性



目标: MAX 支撑距离 d
求解分离平面

SVM求解 · 问题数学化



已知: x_i 特征向量

y_i 标记 (+1 正例, -1 负例)

求解: $\vec{w} \cdot \vec{x} + b = 0$

$\vec{w}$ 法向量, b 截距

要求:

正支撑 $wx + b = 1$

负支撑 $wx + b = -1$

中间 $wx + b = 0$

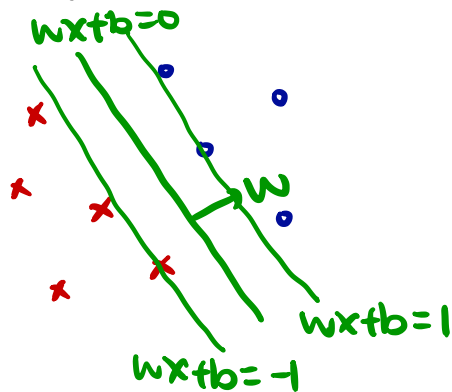
目标: $\text{MAX } d = \frac{2}{\|\vec{w}\|}$

即 $\text{MIN } \frac{\|\vec{w}\|^2}{2}$

$$w(x_1 - x_2) = 2$$

$$\|\vec{w}\| \times d = 2$$

SVM求解 · 问题数字化 CONT'



限制 正例 $w x_i + b \geq 1$

负例 $w x_i + b \leq -1$

统一为 $y_i (w x_i + b) \geq 1$

已知 x_i, y_i

求 w, b

目标, $\min \frac{\|w\|^2}{2}$

限制 $y_i (w x_i + b) \geq 1$

SVM

目标问题



SVM求解 · 线性 · 拉格朗日乘子法

KKT条件: 解决不等式

拉格朗日函数

$$L(w, b, \alpha) = \frac{1}{2} \|w\|^2 - \sum \alpha_i [y_i (w^T x_i + b) - 1]$$

↑ ↑
 拉格朗日乘子

令 $\frac{\partial L}{\partial w} = 0$, $\frac{\partial L}{\partial b} = 0$, 可得 $w = \sum \alpha_i y_i x_i$; $\sum \alpha_i y_i = 0$

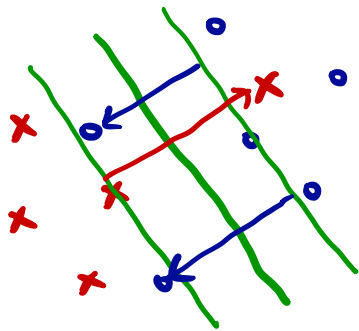
回带得

MAX

$$L(w, b, \alpha) = \sum \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j x_i^T x_j$$

s.t. $\sum \alpha_i y_i = 0$ 支持向量的 $\alpha_i \neq 0$.

SVM求解 · 线性不可分



限制 $y_i(w \cdot x_i + b) \geq 1 - \xi_i$

目标, $\text{MIN } \frac{1}{2} \|w\|^2 + C \sum \xi_i \hookrightarrow$

目标, 惩罚项

拉格朗日乘子法

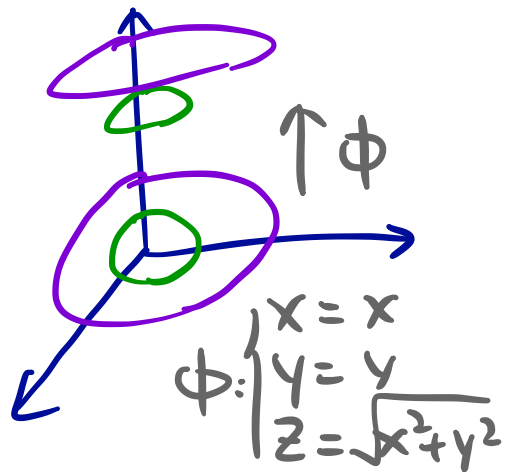
$$\text{MAX } \sum \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j x_i^T x_j$$

s.t. $C \geq \alpha_i \geq 0$ 仅3-4限制

$$\sum \alpha_i y_i = 0$$

SVM求解 · 非线性求解

思路 低维 映射 高维
(非线性) (线性可分)



问题 找到 ϕ

$$\text{MAX } \sum \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j Y_i Y_j$$

$$\phi(x_i)^T \phi(x_j) \\ \uparrow \\ K(x_i, x_j)$$

核函数

$$K(x, z) = \phi(x)^T \phi(z)$$

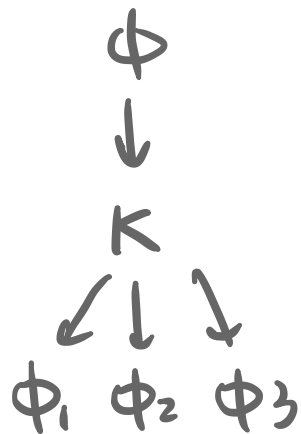
(核函数)

(映射函数)

试核函数

高斯核函数

SVM求解. 非线性



判别 找出的 $wx+b=0$
是在高维的分割平面

对于判定点 x , 应代入 $\phi(x)$.
但 ϕ 很难去找.

核函数优点,

无需放到高维空间

中计算, 无需关心具体的 ϕ

解决 $w\phi(x)+b=0$

$$= (\sum \alpha_i y_i \phi(x_i))^T \phi(x) + b$$

$$= \sum \alpha_i y_i \phi(x_i)^T \phi(x) + b$$

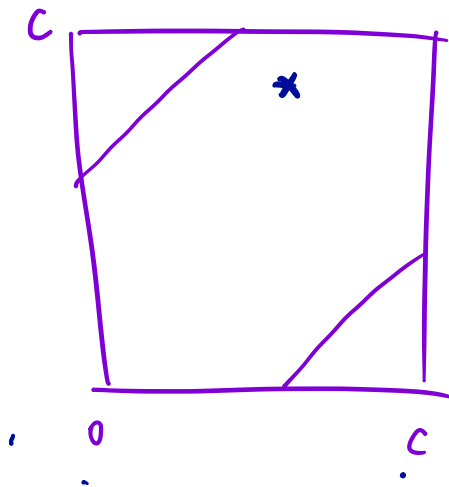
$$= \sum \alpha_i y_i K(x_i, x) + b$$

> 核函数 <

SVM求解 · SMO算法.

目的 求解 $\text{MAX } \sum \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j x_i^T x_j$

$$\text{s.t. } C \leq \alpha_i \leq 0, \sum \alpha_i y_i = 0$$



$$\alpha_i - \alpha_j$$

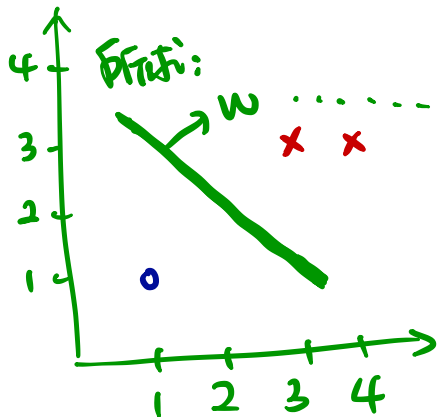
SVM实战

1 不用拉格朗日·硬算

正例1: $x_1 = (3, 3)$, $x_2 = (4, 3)$

负例1: $x_3 = (1, 1)$

求最大间隔超平面



设: $w = (w_1, w_2)$, b

$$wx_1 + b = 1$$

$$wx_3 + b = -1$$

从图可得

$$\text{MIN } w_1^2 + w_2^2$$

$$\begin{cases} 3w_1 + 3w_2 + b = 1 \\ w_1 + w_2 + b = -1 \end{cases}$$

直观有 $w_1 = w_2$

$$\text{得 } w_1 + w_2 = \frac{1}{2}, b = -2$$

$$\text{故 } w = (\frac{1}{2}, \frac{1}{2})^T$$

$$\text{所求平面: } wx - 2 = 0$$

SVM 实战

≥ 拉格朗日算爆!

正例1: $x_1 = (3, 3)$, $x_2 = (4, 3)$

负例1: $x_3 = (1, 1)$.

目标:

$$\text{MAX } \sum \alpha_i - \frac{1}{2} \sum \alpha_i \alpha_j y_i y_j x_i^T x_j$$

$$\text{s.t. } \alpha_i \geq 0, \sum \alpha_i y_i = 0$$

代入

$$\text{MAX } \alpha_1 + \alpha_2 + \alpha_3 - \frac{1}{2} (18\alpha_1^2 + 25\alpha_2^2 + 2\alpha_3^2 + 42\alpha_1\alpha_3 - 12\alpha_1\alpha_2 - 14\alpha_2\alpha_3)$$

$$\text{s.t. } \alpha_i \geq 0, \alpha_1 + \alpha_2 - \alpha_3 = 0$$

将 $\alpha_3 = \alpha_1 + \alpha_2$ 代入

$$S(\alpha_1, \alpha_2) = 2\alpha_1 + 2\alpha_2$$

$$- 4\alpha_1^2 - \frac{13}{2}\alpha_2^2 - 10\alpha_1\alpha_2$$

$$\text{令 } \frac{\partial S}{\partial \alpha_1} = \frac{\partial S}{\partial \alpha_2} = 0 \text{ 知在 } (\frac{3}{2}, -1) \text{ 处取值}$$

但不满足 $\alpha_i \geq 0$.

取 $\alpha_1 = 0$, 则有 $\alpha_2 = \frac{2}{13}$, ...

$\alpha_2 = 0$ 则有 $\alpha_1 = \frac{1}{4}$, ...

SVM实战

2 拉格朗日(续)

$$\text{故 } \alpha_1 = \frac{1}{4}, \alpha_2 = 0, \alpha_3 = \frac{1}{4}.$$

↓
非支撑向量 $\alpha_i = 0$

即找到一支持
↓ 向量

$$\text{用 } w = \sum \alpha_i y_i x_i, b = y_j - \sum \alpha_i y_i (x_i^T x_j) \text{ 其中 } (\alpha_i > 0)$$

$$\begin{aligned} w &= \frac{1}{4}(3, 3) + \frac{1}{4}(1, 1) \times (-1) \\ &= (\frac{1}{2}, \frac{1}{2}). \end{aligned}$$

$$b = -2$$

余下是一样的。

SVM实战

3 非线性!

反例: $x_1 = 0$

正例: $x_2 = 1, x_3 = -1$

核函数: $k(x, z) = (x \cdot z + 1)^2$

目标:

$$\text{MAX } \sum \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j k(x_i, x_j)$$

$$\text{s.t. } \alpha_1 = \alpha_2 + \alpha_3, \alpha_i \geq 0$$

代入

$$\text{MAX } \alpha_1 + \alpha_2 + \alpha_3 - \frac{1}{2} \left(\alpha_1^2 - 2\alpha_1\alpha_2 - 2\alpha_1\alpha_3 + 4\alpha_2^2 + 4\alpha_3^2 \right)$$

$$\text{MAX } 2\alpha_2 + 2\alpha_3 - \frac{3}{2}\alpha_2^2 - \frac{3}{2}\alpha_3^2 + \alpha_2\alpha_3$$

偏导求极值

$$\alpha_1 = 2, \alpha_2 = \alpha_3 = 1$$

求 w 与 b

$$b = -1$$

决策函数

$$f(x) = \text{sign}(2(-1) \cdot 1 + 1 \cdot 1 \cdot (x+1)^2 + 1 \cdot 1 \cdot (1-x)^2 - 1)$$

$$= \text{sign}((x+1)^2 + (1-x)^2 - 3)$$

↑
分隔超平面方程 "0"

SVM实战

4 画点分布

$\alpha_i^* > 0$ 所对应的样本 x_i 称为支持向量（软间隔的支持向量）

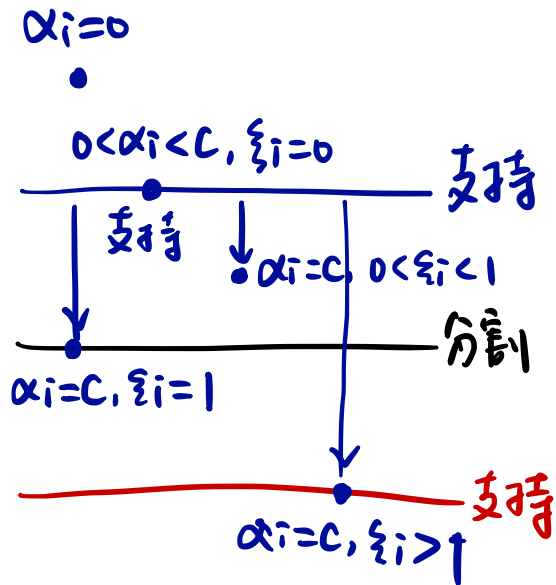
若 $\alpha_i^* < C$, 则 $\xi_i = 0$, x_i 在间隔边界上

若 $\alpha_i^* = C$, $0 < \xi_i < 1$, 则分类正确, x_i 在

间隔边界与分离超平面之间

若 $\alpha_i^* = C$, $\xi_i = 1$, 则 x_i 在超平面上

若 $\alpha_i^* = C$, $\xi_i > 1$, 则 x_i 位于误分一侧



SVM实战

4通点分布

x_1 : $0 < a < C$, $\xi = 0$ 正例1

x_2 : $0 < a < C$, $\xi = 0$ 负例1

x_3 : $a = C$, $\xi = 0.4$ 正例1

x_4 : $a = C$, $\xi = 0.7$ 负例1

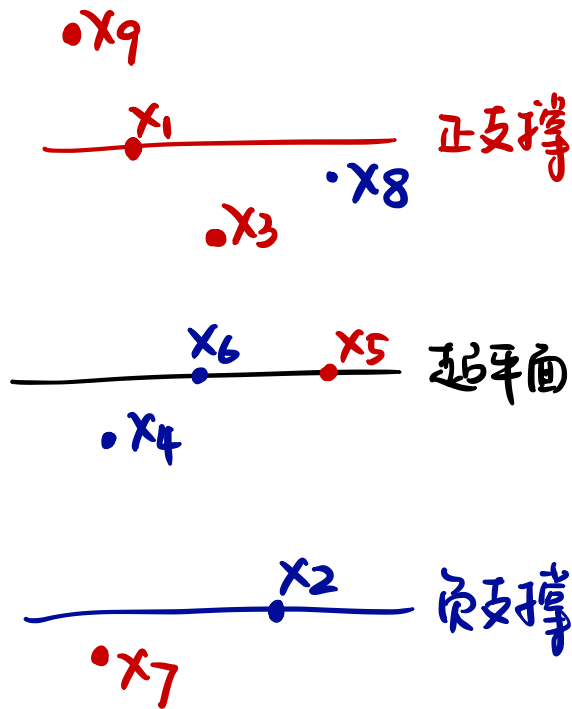
x_5 : $a = C$, $\xi = 1$ 正例1

x_6 : $a = C$, $\xi = 1$ 负例1

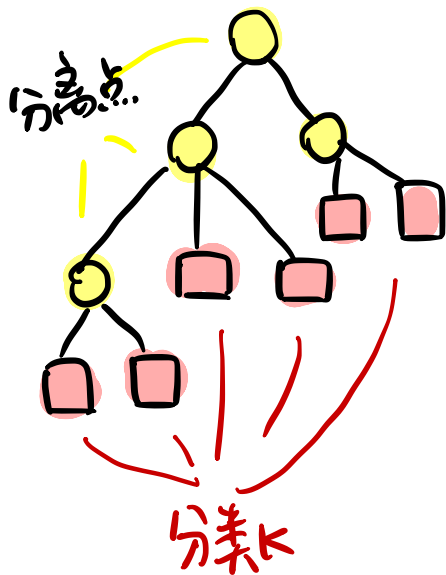
x_7 : $a = C$, $\xi = 2.1$ 正例1

x_8 : $a = C$, $\xi = 1.8$ 负例1

x_9 : $a = 0$, 正例1



决策树



特征选择 每一层以什么特征来分割

└ 信息增益 $H(D) = -\sum \frac{C_k}{D} \log \frac{C_k}{D}$

$$H(D|A) = -\sum \frac{D_i}{D} \sum \frac{D_{ik}}{D_i} \log \frac{D_{ik}}{D_i}$$

$$g(D, A) = H(D) - H(D|A)$$

H代表分类的熵. 越小越整齐.

$$\text{故 } \min H(D|A) \rightarrow \max g(D, A)$$

└ 信息增益比 $H_A(D) = -\sum \frac{D_i}{D} \log \frac{D_i}{D}$

$$g_R(D, A) = \frac{g(D, A)}{H_A(D)} \text{ 优化.}$$

决策树

ID3算法

特判 1 同一类 $C_k \rightarrow C_k$
2 无特征 $\rightarrow \text{MAX-CNT} \cdot C_k$

选特征 3 $\text{MAX } g(D, A)$ 的 $A_g \leftarrow$

预剪枝 4 $A_g < \text{阈值} \rightarrow \text{MAX-CNT} \cdot C_k$

分割 5 按照 A_g 进行分割 D , 并分配

特判空 6 若 D 为空 $\rightarrow \text{MAX-CNT} \cdot C_k$

递归 7 递归求解

C4.5算法

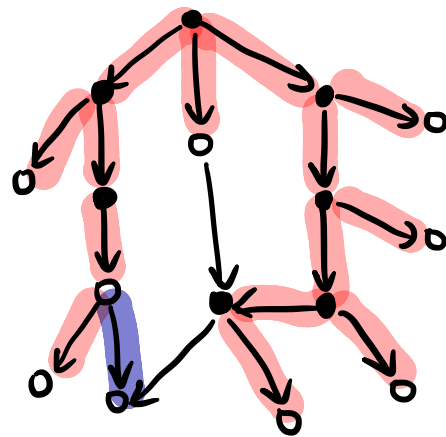
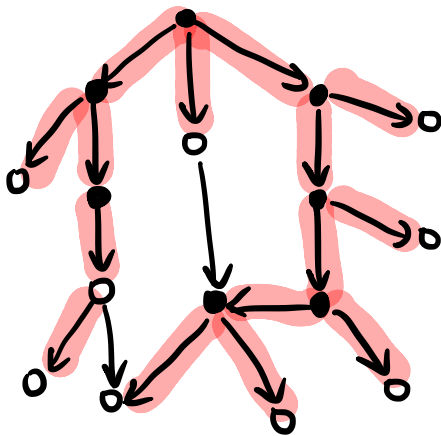
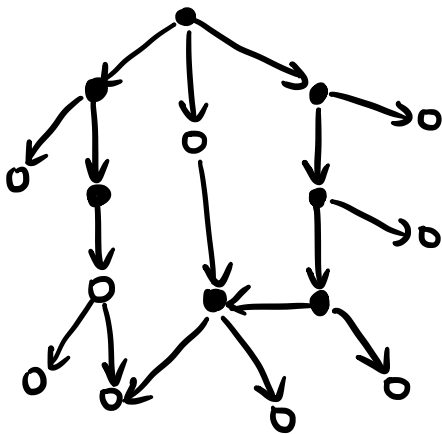
选前 n 的
再选其中 增益比
最大的。

第一章

Open : 准备下一步搜索

closed: 已完成搜索

指针 : 在G中的父亲,



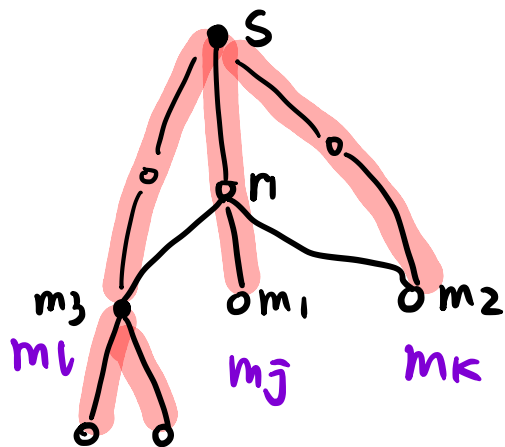
图搜索

三类结点

m_j : 未在open和closed出现过

m_k : open出现

m_l : closed出现



算法流程

1 初始化

2 $n = \text{open}$ 的头

3 如果 n 为目标, 返回

4 扩展 n 的子结点 m_i

5 m_j : 连接 $n \rightarrow m_j$

6 m_k, m_l : 判断 $n \rightarrow m_k, m_l$

7 判断是否修改 m_l 到其后代

8 重排OPEN

图搜索

DFS 深度优先



将 m_j 加入队列:

将 m_j 插入到 open

的最前面



BFS 广度优先



将 m_j 加入队列:

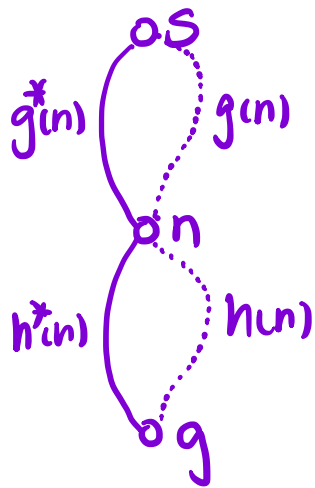
将 m_j 插入到 open

的最后面.



启发式搜索

评价函数 F



$f^*(n)$ 确实的 $s-n-g$ 最短路径

$f(n)$ 估计的最短路径

算法流程

1 初始化

2 $n = \text{OPEN}$ 的头

3 如果 n 为目标, 返回

4 扩展 n 的子结点 m_i

$$f(n, m_i) = g(n, m_i) + h(m_i)$$

5 m_j : 连接 $n \rightarrow m_j$

6 m_k : $f(n, m_k) < f(m_k)$
标记 $n \rightarrow m_k$. 并更新 m_k

7 m_l : $f(n, m_l) < f(m_l)$
标记 $n \rightarrow m_l$. 更新 m_l .

并重新插入 OPEN.

8 按照 f 的大小重排 OPEN

A* 算法.

$h(n) \leq h^*(n) \rightarrow$ 满足这个就是 A* 算法

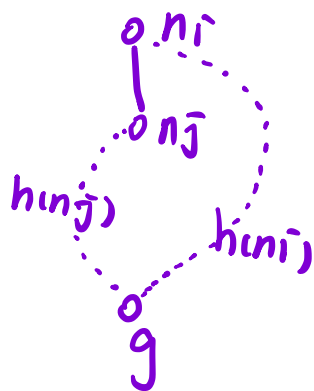
$h^*(n)$ 从节点 n 到目标 g 的最短路径耗数值

$h(n)$ 对 $h^*(n)$ 的估计值. 启发函数

极端: $h(n) = 0$. | 但应尽量接近 h^* 的大.

性质1 h 单调: $h(n_i) - h(n_j) \leq C(n_i, n_j)$
 $h(n_i) \leq C(n_i, n_j) + h(n_j)$

结论1 若 h 单调, A* 对扩展 n 后, 有 $g(n) = g^*(n)$



A* 算法改进

性质2

OPEN上, $f(n) < f^*(s)$ 肯定会被A*扩展

A* 选作扩展, 肯定有 $f(n) \leq f^*(s)$

结论2

OPEN = (NEST |) $\leftarrow f^*(s)$

NEST = { OPEN中满足 $f < f_m$ 的节点集合 }

2

if NEST $\neq \emptyset$ THEN $n = \min(g_i)$

ELSE $n = \text{open第-1}, f_m = f(n).$

A* 算法改进

性质3

若 A_2 比 A_1 有更多的启发信息, $h_2(n) > h_1(n) \quad \forall n$
则 A_2 扩展的 A_1 均扩展过了.

结论3

h 越接近 h^* 效果越好.

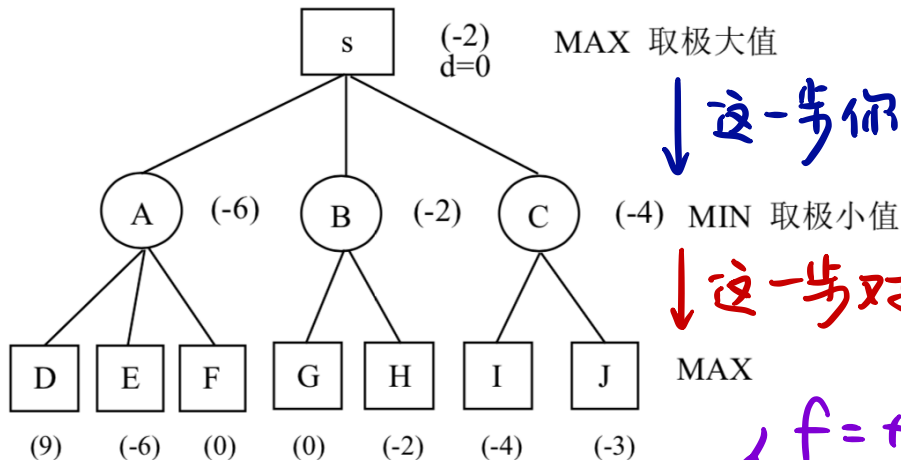
! $h_2(n) \geq h_1(n)$ 时结论则不成立

A* 若干题目

- (1) 设计A*的h函数, 迷宫问题
- (2) 画出A*搜索树, 给出最优解. (顺序 / f值 / 解路径)
- (3) h单调的条件! 目的
- (4) NEST的条件! 目的. $f^*(s)$
- (5) $h_1 \geq h_2$. 不保证 节点1 $\leq$ 节点2. 给出例子

极大极小搜索 (与或图搜索)

第二章



↓ 这一步你走, 取得分最大

↓ 这一步对手走, 取得分最小。

$f = +\infty$ 我胜利
 $f = -\infty$ 对手胜利

算法流程:

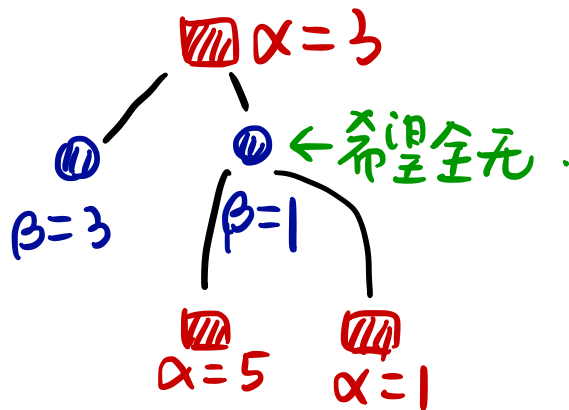
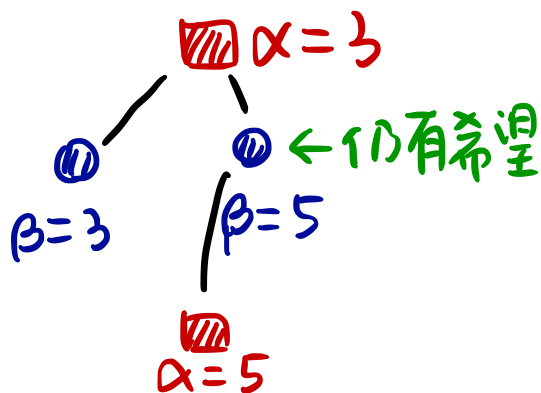
- 1 宽搜所有的点
- 2 由低往高返回计算值

α - β 搜索

α : 极大节点, MAX的下界
 β : 极小节点, MIN的上界 } 过程中的, 可能会继续变^大_小

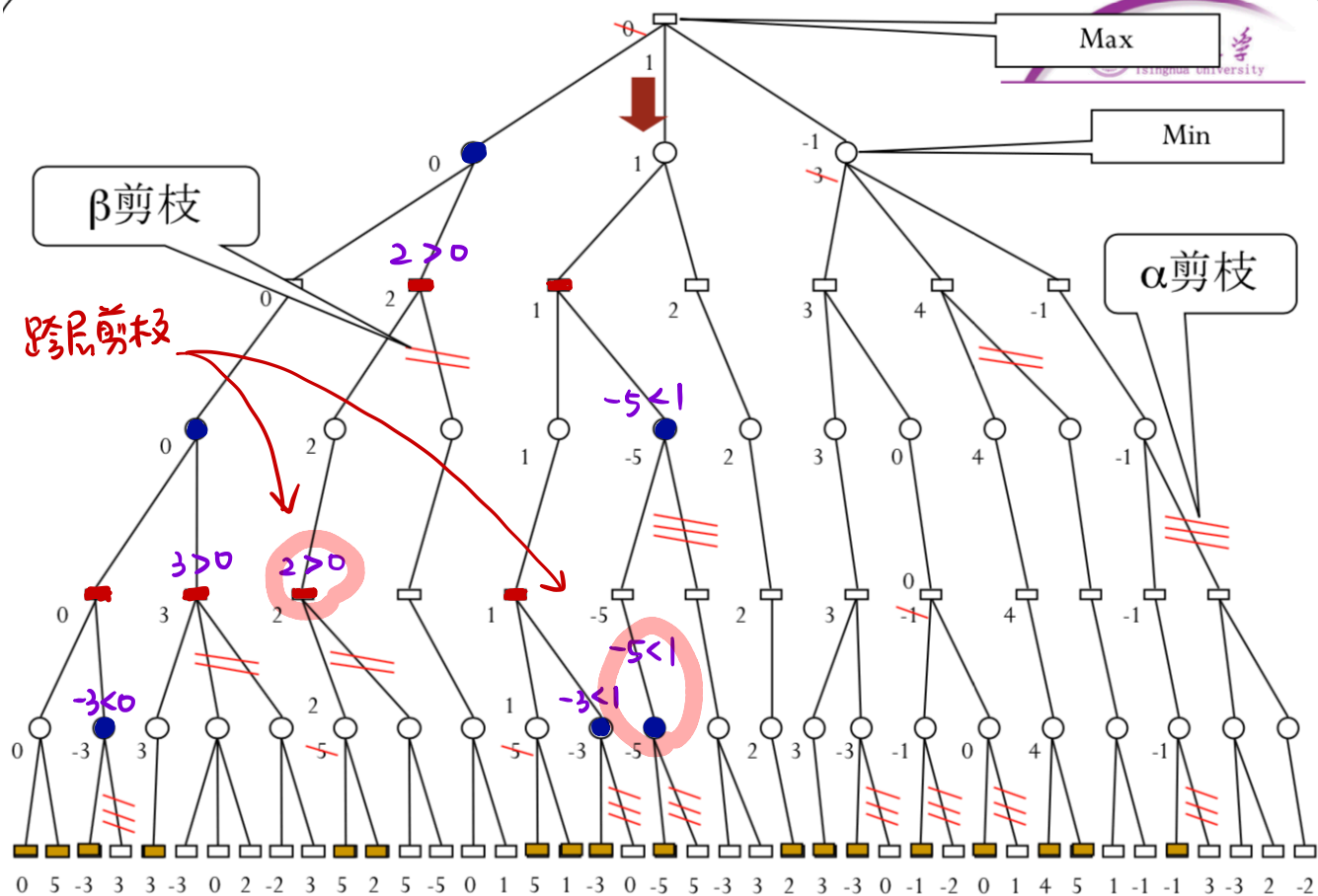
剪枝条件: 后辈 $\beta \leq$ 祖先 α . 剪!

后辈 $\alpha \geq$ 祖先 β . 剪!



α - β 搜索

注意跨层剪裁!!



信心上限算法 UCB、信心上限树 UCT

四子棋大家写过。

高级搜索<第3章>

局部搜索

模拟退火

遗传算法

应该不考大题!

填空题 / 判断题

问答题吧!

神经网络与

深度学习

第五章



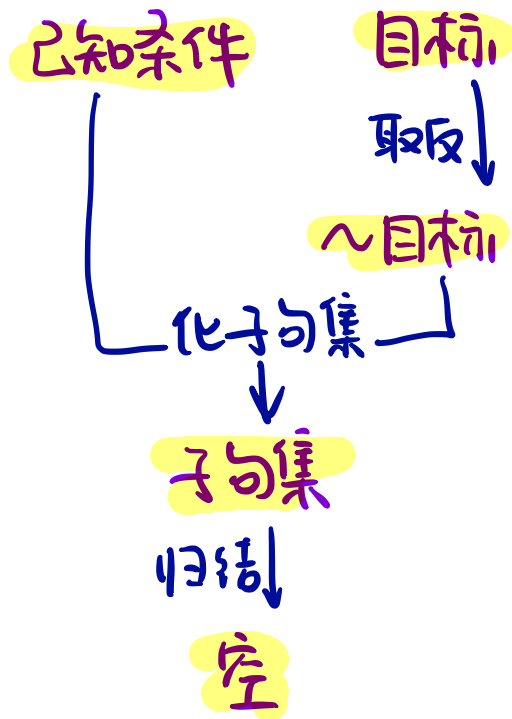
分析神经元的

...

求偏导!

谓词逻辑 第六章

归结



化子句集

1 消 $\rightarrow$ 和 $\leftrightarrow$

$$a \rightarrow b$$

$$\sim a \vee b$$

2 $\sim$ 内移

$$\sim(a \vee b)$$

$$\sim a \wedge \sim b$$

$$\sim(a \wedge b)$$

$$\sim a \vee \sim b$$

$$\sim(\exists x)P(x)$$

$$(\forall x) \sim P(x)$$

$$\sim(\forall x)P(x)$$

$$(\exists x) \sim P(x)$$

3 变量标准化

$$(\exists x)A(x) \vee (\exists x)B(x)$$

$$(\exists x)A(x) \vee (\exists y)B(y)$$

4 量词左移

$$(\exists x)A(x) \vee (\exists y)B(y)$$

$$(\exists x)(\exists y) A(x) \vee B(y)$$

5 消存在

$$(\exists z)(\forall x)(\exists y) x, y, z$$

$$\forall x, x, f(x). a$$

$\downarrow$ $\downarrow$
 y z

化子句集

6 化为合取范式

化成 $(a \vee b) \wedge (c \vee d) \wedge (e \vee f)$

7 隐去全称量词

$\forall(x) P(x)$ $P(x)$

8 表示为子句集

化成 $a \vee b, c \vee d, e \vee f$

9 变量标准化

归结方法

$$C_1 = L \vee C_1'$$

$$C_2 = \neg L \vee C_2'$$

$$\Rightarrow C = C_1' \vee C_2'$$

置换

例子

$$S = \{z/x, A/y\}$$

$$p[x, f(y), B]_S = p[z, f(A), B]$$

合一

$$E_{1S} = E_{2S} = \dots = E_{nS}, \text{ 称 } S \text{ 为 } E_i \text{ 的合一者}$$

$$p(x, f(y), B), p(z, f(B), B)$$

$$S = \{z/x, B/y\} \text{ 为合一者}$$

最一般合一者 (mgu)

置换最少, 限制最少, 产生的最有一般性.

$$p(x, x, z)$$

$$p(f(y), f(B), y)$$

$$S = \{B/z, B/y, f(B)/x\}$$

命题的归结树

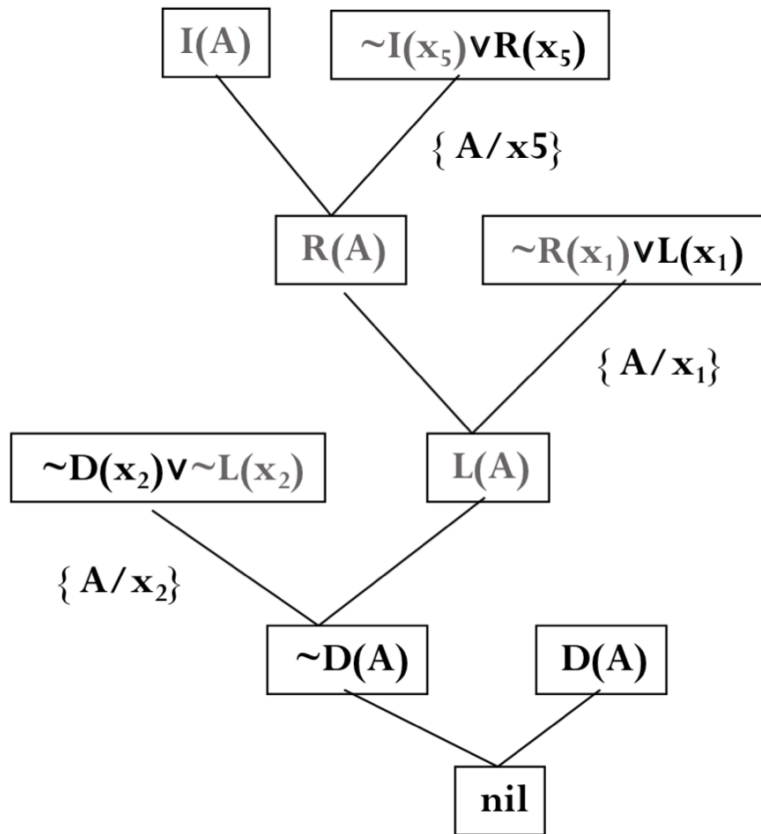
$\sim R(x_1) \vee L(x_1)$

$\sim D(x_2) \vee \sim L(x_2)$

$D(A)$

$I(A)$

$\sim I(x_5) \vee R(x_5)$



提取答案, 修改证明树

子句集:

$$\sim \text{AT}(\text{John}, x_1) \vee \text{AT}(\text{Fido}, x_1)$$
$$\text{AT}(\text{John}, \text{School})$$
$$\sim \text{AT}(\text{Fido}, x_2)$$

重言式代替归结求反的句

