人工智能导论期末复习

Chap 1. 搜索问题

一、盲目搜索

1. 深搜（return 路径或 Fail）

判断 ⟨ 到目标， return null.
走到头． Fail.

⇓

应用规则获得可走集合，若为空， return Fail.

⇓

取第一个继续走 （…）

2. 宽搜（G - 搜索图， OPEN - 队列， CLOSED - 已搜过）
（OPEN表按深度从小到大排序）

二、启发式图搜索

1.* A算法（OPEN表按f值从小到大排序），注意扩展n时要更新有父的 m_k, m_l (OPEN) ADD ↑ (CLOSED)

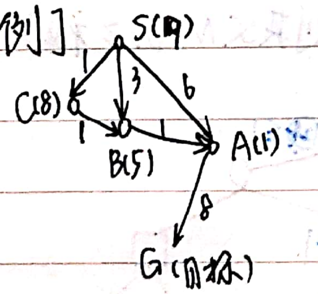
☆结束条件：OPEN表取出第一个，为goal.

2.* A*： $h(n) \le h^*(n)$（真实）， 改进 $h(n) - h(m_i) \le c(n, m_i)$ （$m_i \in EXPAND(n)$）

- 改进A*的 f_m —— $\max f(n_i)$， $n_i \in CLOSED$

利用 f_m 划分出 OPEN的子集 NEST 为 $f(n_i) < f_m$ 的 $\{n_i\}$，优先其中 g 最小的

[例]



OPEN表	CLOSED表	f_m
S(0+9)	S(0+9)	9
C(1+8), B(3+5), A(6+1)	S(9), ~~A(7)~~ B(3+5)	9
C(1+8), A(4+1)	S(9), B(8), A(4+1)	9
C(1+8), G(12+0)	S(9), B(8), A(5), C(1+8)	9
B(2+5), G(12+0)	S(9), B(8), A(5), C(9), B(2+5)	9
A(3+1), G(12+0)	S, B_8, A_5, C_9, B_7, A(3+1)	9
G(11+0)	······ A_4, G(11+0)	9

注意是选 g_{min}

NEST空，n = OPEN 1st
$f_m = f(n) = 9$

end. □

三、动态规划（$h \equiv 0$）

viterbi算法：遍历parent可确定从根经p到v的最优 P_v.
（到p为gp）

Chap 2. 对抗搜索

一. * α-β 剪枝

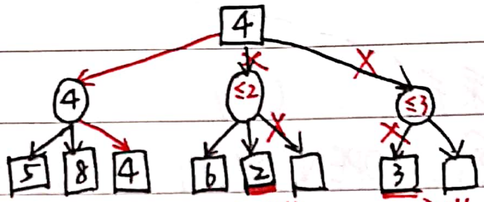
极大：己方选择，必选分数极大的；极小：对方选择分数极小的（己方）（相对于自己）

△ 走步标记：只能走一步

剪枝：α剪枝即MAX节点的孙子≤自己时，孙子的父亲被剪。（β≤α）

β剪枝即MIN 孙子≥自己

[例1] MAX MIN MAX



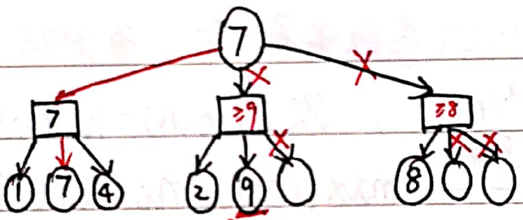
(α剪枝)

(MAX的下界≤祖父，则其父MIN≤祖父，必不选)

2≤4，3<4

评估局面变差

MIN MAX MIN

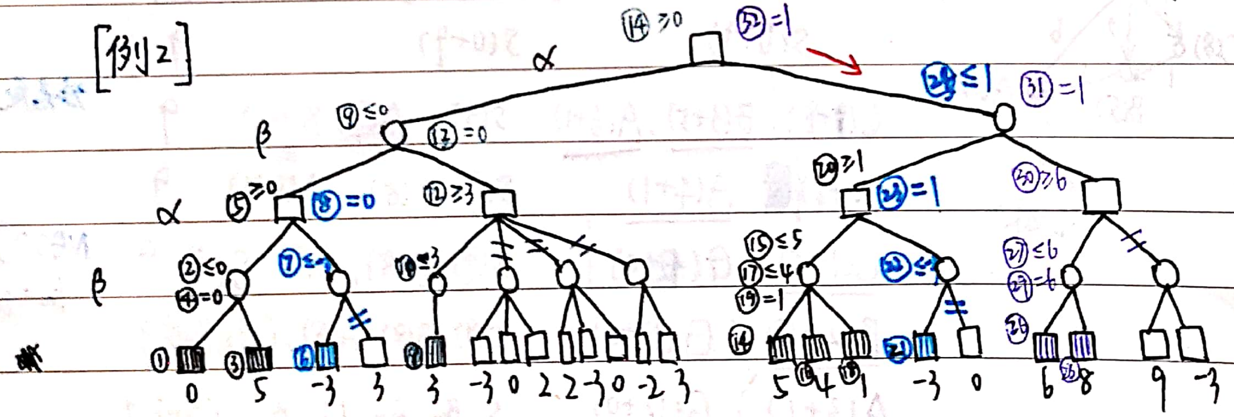


(β剪枝)

9≥7

有利局面减弱 (MIN的上界超过祖父，则其父MAX≥祖父，必不被选)

[例2]



二. 蒙特卡罗树搜索

区别：MCTS随机落子，UCT按信心上限选（先扩展，不可扩展就BestChild）

（$\bar{X}_j$ 平均收益，$I_j=\bar{X}_j+c\sqrt{\frac{2\ln n}{T_j(n)}}$（访问次数占比））

Chap 3. 高级搜索

一. 局部搜索（～梯度下降）

问题：
- 局部最优（收敛到 local max/min 但非全局） ⇒ 按概率选点（f优P大）
- 步长（learning rate 大会难收敛） ⇒ 变步长
- 起始点 ⇒ 随机生成多个起始点，优中选优

二.* 模拟退火

1. 状态转移

① 低能方向（好的点）：一定转移 （Random < P_0，转移✓）

② 高能方向（差的点）：按概率转移，接受概率 $P_0 = e^{-\frac{\Delta E}{KT}}$ （$\Delta E > 0$）
T-温度
K-玻常数

2. 初始温度 $t_0 = \dfrac{\Delta f(i,j)}{\ln(P_0^{-1})}$，$P_0$为给定较大的接受概率，

估计 $\Delta f(i,j) = \max\limits_{i\in S} f(i) - \min\limits_{j\in S} f(j)$，S为随机产生的序列

温度下降：等比例↓，$t_{k+1} = \alpha t_k$，$\alpha \in (0,1)$.

结束条件：
- 温度T_k下平衡条件：固定长度 L_k ?
- 算法终止：零度法，设定 $\varepsilon > 0$，$t_k < \varepsilon$ 时 end.

[例] a, b, c, d 依次产生的状态，f(a)=10，f(b)=20，f(c)=30，f(d)=20.
算法循环3次，t=10 不变，随机数依次为 0.5，0.1，0.01，0.2，求解.

次数	i	j	f(i)	f(j)	P	Random	
①	a	b	10	20	$e^{-\frac{10}{10}} = 0.37$	0.5	×
②	a	c	10	30	$e^{-\frac{20}{10}} = 0.14$	0.1	✓
③	c	d	30	20	1 （低能✓）	0.01	✓

最后的解为 d.

三、遗传算法

1、编码：数字取值有16个，可以用4位二进制数；
（矩阵展开）　　　　也可以用十进制.

2. 选择：轮盘赌法，按概率 $p(X_i)=\frac{F(X_i)}{\sum_{j=1} F(X_j)}$ 选择

交叉：二进制时，双亲在交叉位置后的完全对换

变异：二……，变异位置 0→1 / 1→0 .

3. 适应函数 —（一般）问题指标函数 $f(x)$

（f_m 周围平缓）加速 — 非线性 $f_1(x)=\begin{cases}\frac{1}{f_{max}-f(x)}, & f(x)<f_{max}\\ M, & else\end{cases}$

线性 $f_2(x)=\alpha f(x)+\beta$

△ 十进制交叉：2P ⇒ 2F，$F_{1(2)}$ 前半取 $P_{1(2)}$，后半按顺序选 $P_{2(1)}$（~~未出现~~ 剩下）的.

变异：2个位置，后一移到前一前面.

4. 结果：全局最优解（每次保留全过程中最好的）

[例] 编码方案：(x-1)（1~16：4位二进制），求 $f(x)=x$ 最大值.（随机数 0.2，0.7，0.67，0.9，0.5）

① 分母是群体适应值之和，即 11+13+4+9=37

个体id	初始群体	编码	适应值	选择概率	选中次数
1	11	1010	11	(0.3) ~~(0.08)~~ $\frac{11}{37}$	1
2	13	1100	13	(0.35) ~~(0.096)~~ $\frac{13}{37}$	0
3	4	0011	4	(0.11) $\frac{4}{37}$	2 可以重复
4	9	1000	9	(0.24) $\frac{9}{37}$	1

（和为N=4）

△ 模拟轮盘赌（把轮盘展开 | ① 0.3 | ② 0.65 | ③ 0.76 | ④ 1 | ）

① r=random=0.2，①√ ⇒ ② r=0.7，③√　③ r=0.67，③√，④ r=0.9，④√

x_b=1100 (13).

△（交叉位是n，表明 $a_1\cdots a_n | a_{n+1}\cdots$）

（新群体）		给定交叉对象	交叉位	子代	子代适应值
1	10|10	2	2	1011	12
2	00|11	1	2	0010	3
3	001|1	4	3	0010	3
4	100|0	3	3	1001	10

x_b=1100(13) 不变，解答的编码为1100，解码后 x=13.

Chap 4. 统计机器学习

一. 朴素贝叶斯

1. 训练/学习过程就是统计 & 求概率.

① 各类概率 $P(Y=c_k)=\dfrac{\sum_{i=1}^{N} I(y_i=c_k)+\lambda}{N+K\lambda}$ (K个类), 频率决定概率. (Laplace平滑)

② 条件概率

$$P(X^{(j)}=a_{jl}\mid Y=c_k)=\frac{\sum I(x_i^{(j)}=a_{jl},\ y_i=c_k)+\lambda}{\sum I(y_i=c_k)+S_j\lambda},\quad \frac{\text{有特征}a_{jl}\text{的}c_k\text{类数量}}{c_k\text{类总数}}$$

2. 分类时的概率 (经归一化, 去分母)

$$P_k'(Y=c_k\mid X=x)=P(Y=c_k)\cdot\prod_{j=1}^{n}P(X^{(j)}=x^{(j)}\mid Y=c_k),\quad y=\arg\max_{c_k} P_k'.$$

(条件已知特征为X) (认为每个特征独立, 相乘计算)

二. 支持向量机 (SVM)

1. 线性可分SVM ① 内积

对偶问题: $\min_{\alpha}\left\{\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j (x_i\cdot x_j)-\sum_{i=1}^{N}\alpha_i\right\}$, s.t. $\sum_{i=1}^{N}\alpha_i y_i=0$, $\alpha_i\ge 0$

$\Rightarrow w^*=\sum_{i=1}^{N}\alpha_i^* y_i x_i$, $b^*=y_j-\sum_{i=1}^{N}\alpha_i^* y_i(x_i\cdot x_j)$

求min: 偏导为0再求解, 若不满足约束$\alpha_i\ge 0$, 则在边界(某$\alpha=0$)

[例1] $N=3$, 正例 $x_1=(1,0)^T$, 负例 $x_2=(4,3)^T$, $x_3=(2,1)^T$, 求最大间隔超平面

对偶得 $\min_{\alpha}\frac{1}{2}(\alpha_1^2+25\alpha_2^2+5\alpha_3^2-8\alpha_1\alpha_2-4\alpha_1\alpha_3+22\alpha_2\alpha_3)-\sum\alpha_i$

~~min~~ s.t. $\alpha_1-\alpha_2-\alpha_3=0$, 代入 $\alpha_1=\alpha_2+\alpha_3$ 得

$S(\alpha_2,\alpha_3)=\cancel{\frac{1}{2}(26\alpha_2^2+6\alpha_3^2-8\alpha_2^2-8\alpha_2\alpha_3}\ \frac{1}{2}\left((\alpha_2+\alpha_3)^2-4(\alpha_2+\alpha_3)(2\alpha_2+\alpha_3)+\cdots\right)$

$=\frac{1}{2}(\alpha_2^2+2\alpha_2\alpha_3+\cancel{3}\alpha_3^2-8\alpha_2^2-4\alpha_3^2-12\alpha_2\alpha_3+25\alpha_2^2+22\alpha_2\alpha_3+5\alpha_3^2)-2\alpha_2-2\alpha_3$

$=9\alpha_2^2+\cancel{4}\alpha_3^2+6\alpha_2\alpha_3-2\alpha_2-2\alpha_3$

令偏导为0得 $\begin{cases}\frac{\partial S}{\partial\alpha_2}=18\alpha_2+6\alpha_3-2=0\\ \frac{\partial S}{\partial\alpha_3}=\cancel{4}\,2\alpha_3+6\alpha_2-2=0\end{cases}\Rightarrow \cancel{\begin{cases}\alpha_2=-\frac{1}{9}\\ \alpha_3=\frac{2}{3}\end{cases}}$ × 在边界 ② 无解

$\alpha_2=0$ 时, $\min S(\alpha_3)=S(0,\cancel{\frac{1}{2}}\,1)=\cancel{-\frac{1}{2}}\,-1$ √

$\alpha_3=0$ 时, $(2\alpha_3^2-2\alpha_3)$ $S(\frac{1}{9},0)=-\frac{1}{9}$

$\Rightarrow \alpha_1=0+\cancel{\frac{1}{2}}\,1=\cancel{\frac{1}{2}}\,1$, $\alpha_2=0$, $\alpha_3=\cancel{\frac{1}{2}}\,1$, x_1, x_3 为支持向量.

$w^*=\cancel{\frac{1}{2}}(x_1-x_3)=(\cancel{-\frac{1}{2}}\,-1,\ \cancel{-\frac{1}{2}}\,-1)$, $b^*=1-(\cancel{\frac{1}{2}}x_1^2-\cancel{\frac{1}{2}}(x_1\cdot x_3))=1-(\cancel{\frac{1}{2}}\,1-\cancel{1}\,2)=\cancel{\frac{3}{2}}\,2$.

分离超平面为 $w^*x+b^*=0$, 即 $-\cancel{\frac{1}{2}}x^{(1)}-\cancel{\frac{1}{2}}x^{(3)}+\cancel{\frac{3}{2}}\,2=0$

分类决策函数 $f(x)=\mathrm{sign}(-\cancel{\frac{1}{2}}x^{(1)}-\cancel{\frac{1}{2}}x^{(3)}+\cancel{\frac{3}{2}})$.

2、~~非线性SVM~~ 线性SVM(某些点不可分，引入松弛变量 ξ_i)

对偶问题 $\min_\alpha$ 相同，s.t. $\sum \alpha_i y_i = 0$，$0 \le \alpha_i \le C$

求得最优解后计算 w^*(同)，选择 $\alpha_j^* \in (0, C)$ 算 b^*.

$\alpha_i^* > 0$，x_i 为支持向量

- $< C$ 则 $\xi_i = 0$，x_i 在边界
- $= C$
  - 若 $\xi_i \in (0,1)$，则分类 ✓
  - $\xi_i = 1$，则在超平面上
  - > 1，则误分

3、非线性SVM

· 利用核函数映射为超平面

对偶时内积 $(x_i \cdot x_j)$ 替换为 $K(x_i, x_j)$

三、决策树

熵 $H(X) = -\sum p_i \log p_i$，$p_i = P(X = x_i)$，$H(D)$ - p_i 由数据集D得到

条件熵 $H(Y|X) = \sum p_i H(Y|X = x_i) = \sum_i \frac{|D_i|}{|D|} H(D_i)$　$(p_i = \frac{|C_i|}{|D|})$

$= -\sum_{i=1}^{n} \frac{|D_i|}{|D|} \sum_{k=1}^{K} \frac{|D_{ik}|}{|D_i|} \log \frac{|D_{ik}|}{|D_i|}$.

信息增益(特征A对数据集D)：$g(D, A) = H(D) - H(D|A)$

(全集中算各类 p_i)

(特征子集中算各类 p_i)

按特征划分的子集

1. ID3

先判断：① 不必再分？($All \in C_k$ 类) D中 → return 单节点树T，标 C_k

② 特征不够？(A为空) → 单节点树T，标最多实例的 C_k

③ (信息增益 max < 阈值 ε) → 单　T，最多　C_k

特征有但不明显？

否则 按 A_g 的所有可能分类分成 D_i，

- 若某 D_i 为空(该特征取值下一个实例也无)，标为D中最多的 C_k
- 否则 ~~对~~ 以 D_i 为训练集，$A - \{A_g\}$ 为特征集，递归 (return 子树 T_i)

2. C4.5

根据信息增益比，除以 $H_A(D) = -\sum_{k=1}^{n} \frac{|D_k|}{|D|} \log_2 \frac{|D_k|}{|D|}$，对多分枝特征惩罚 ↑

对连续值属性找到一个分界值 a_0，≤左，>在右.

3. 后剪枝 (验证集上测试剪枝，训练集建出 original DT)

数据少也可直接在 training set 上

Chap 5. 神经网络 & 深度学习

一、神经元

$$o=\sigma\left(\sum_{i=0}^{n} w_i x_i\right)=\sigma(\vec{w}\cdot\vec{x})$$

误差函数（整个训练集） $E(\vec{w})=\frac{1}{2}\sum_{d\in D}\sum_{k\in outputs}(t_{kd}-o_{kd})^2$

（样例）（label－实际）

二、△ BP算法：

$E_d(\vec{w})=\frac{1}{2}\sum_{k\in outputs}(t_k-o_k)^2$ ， $\Delta w_{ji}=-\eta\frac{\partial E_d}{\partial w_{ji}}$ ， $\eta\in(0,1)$

由链式法则， $\frac{\partial E_d}{\partial w_{ji}}=\frac{\partial E_d}{\partial net_j}\cdot\frac{\partial net_j}{\partial w_{ji}}=\frac{\partial E_d}{\partial net_j}x_{ji}$ （ $net_j=\sum_l w_{jl}x_{jl}$ ）

分输出层、隐含层计算 $\frac{\partial E_d}{\partial net_j}$.

三、CNN

激活函数（ReLU 只取≥0）

卷积：对应位置×再＋． ☆注意 $o_j=\text{kernel}\oplus\sigma(net_j)$

△记得＋b！

四、各种 Net（见 byz 笔记）

五、过拟合解决

正则项（L_1. L_2.），Dropout，验证集

六、梯度消失（解决：ReLU / 改善网络结构 {GoogleNet, 残差网络}）

sigmoid 导数 ≤ ¼，层数深了越乘越小．

? 为什么 $|W_{kh}|$ 大会导致 $\sigma_h'\approx 0$ ，过大导致梯度爆炸？

▷

七、word2vec

上下文 MLE，BP求解 $\max_{\theta}\left(\sum_{w\in C}\log\left(\underbrace{F(w,\text{context}(w),\theta)}_{\text{CBOW}+\text{层次化softmax}}\right)\right)$

① CBOW：2C大小的窗口，

求和作为 context，即 $x_w=\sum_{i=1}^{2C}\text{context}(w)_i$

② Huffman树：输入 x_w，输出概率（按每层向左/右做 softmax，累乘）

$\Rightarrow Loss=-\log\prod_{i=2}^{l^w}p(d_i^w\mid x_w,\theta_{i-1}^w)=-\sum_{i=2}^{l^w}\{(1-d_i^w)\log[\sigma(x_w^T\theta_{i-1}^w)]+d_i^w\log[1-\sigma$

（根无01编码） d_i^w：方向（左0右1）

$(1-d_i^w)$：左， d_i^w：右

θ_{i-1}^w：每个节点θ不同

八. RNN

BP求导：交叉熵 + softmax.

(1) 输出层

① $\frac{\partial H_t}{\partial net_j} = \sum_i \frac{\partial H_t}{\partial O_i} \cdot \frac{\partial O_i}{\partial net_j}$ （链式 + Jacobi矩阵）

② $\frac{\partial H_t}{\partial O_i} = -\frac{t_i}{O_i}$, $\boxed{\frac{\partial O_i}{\partial net_j} = \begin{cases} (i=j) & O_j(1-O_j) \\ (i\neq j) & -O_j \cdot O_i \end{cases}}$

$$\Rightarrow \sum_i \frac{\partial H_t}{\partial O_i} \cdot \frac{\partial O_i}{\partial net_j} = \sum_{i\neq j} \left(-\frac{t_i}{O_i}\right) \cdot (-O_j \cdot O_i) + \underset{(i=j)}{\left(-\frac{t_j}{O_j}\right)} \cdot O_j(1-O_j)$$

$$= \sum_{i\neq j} t_i \cdot O_j - t_j(1-O_j) = \sum_{i\neq j} t_i O_j + t_j O_j - t_j$$

$$= \sum_i t_i O_j - t_j = \boxed{O_j \sum_i t_i - t_j}$$

(2) 隐含层

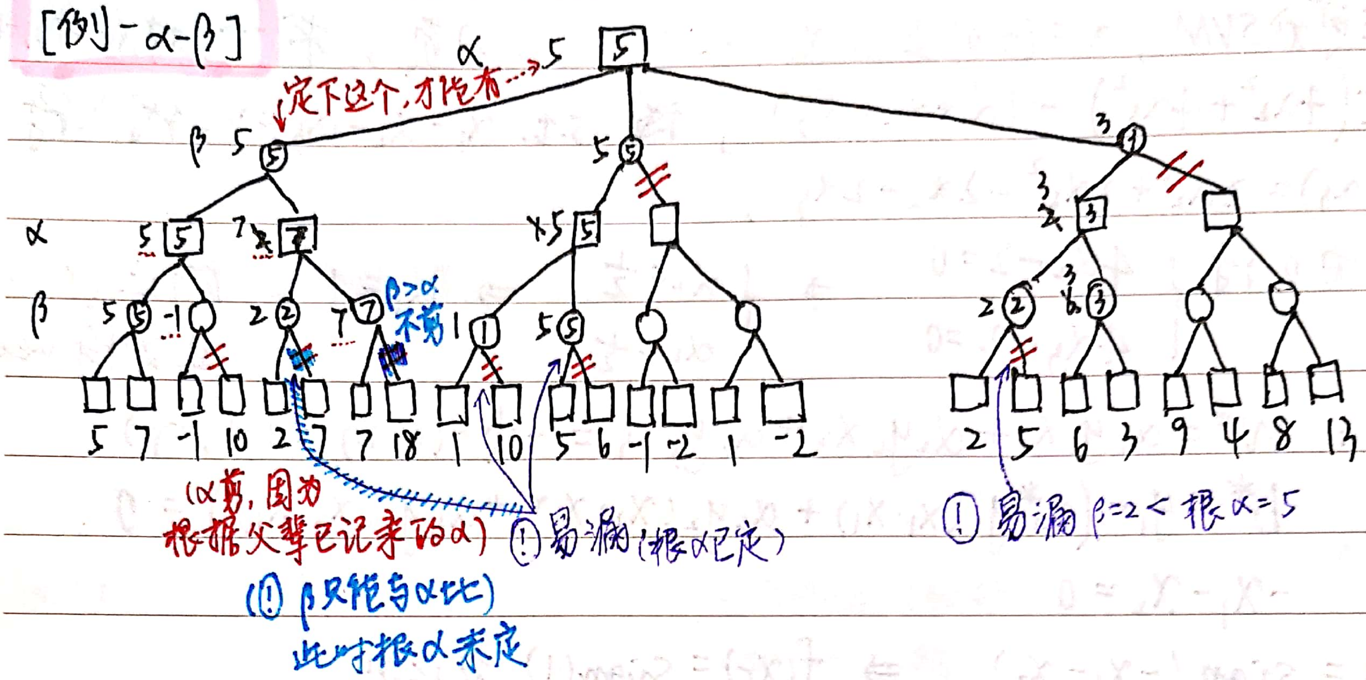
① $\frac{\partial H_t}{\partial net_j} = \sum_{k\in Downstream(j)} \frac{\partial H_t}{\partial net_k} \cdot \frac{\partial net_k}{\partial net_j}$, 而 $\frac{\partial net_k}{\partial net_j} = \sum_i \frac{\partial net_k}{\partial O_i} \cdot \frac{\partial O_i}{\partial net_j}$

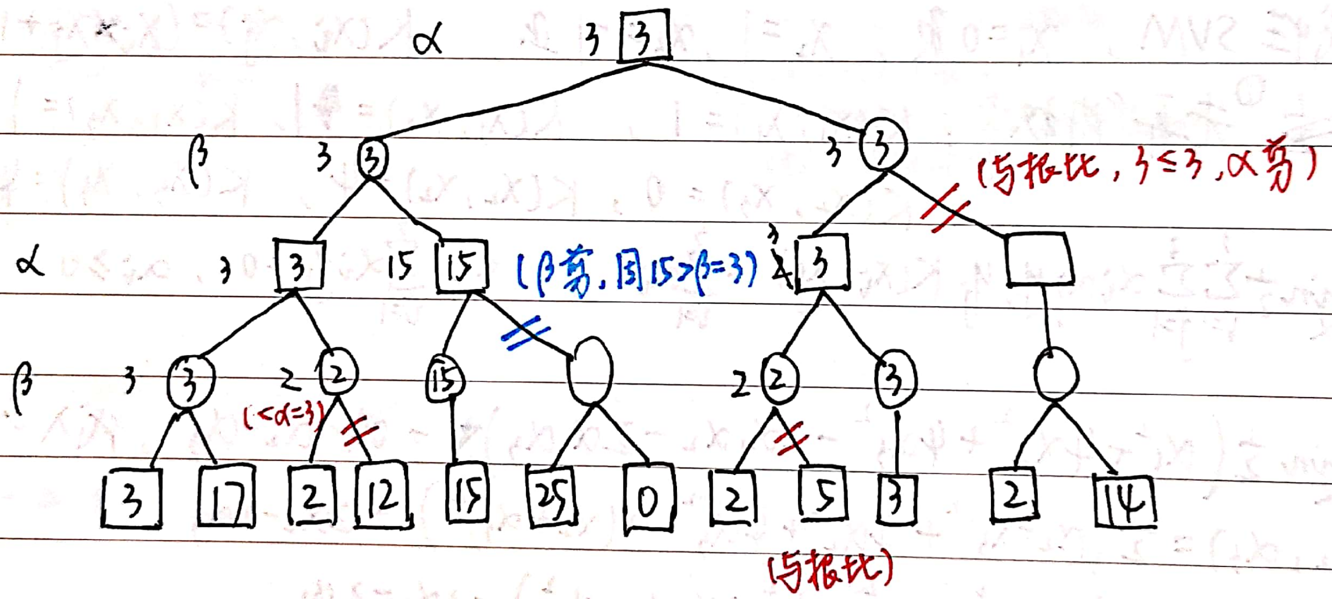
② $\frac{\partial net_k}{\partial O_i} = W_{ki}$, $\frac{\partial O_i}{\partial net_j}$ (见上)

$$\Rightarrow \sum_i \frac{\partial net_k}{\partial O_i} \cdot \frac{\partial O_i}{\partial net_j} = \sum_{i\neq j} W_{ki} \cdot (-O_j O_i) + W_{kj} O_j (1-O_j) = W_{kj} O_j - O_j \sum_i W_{ki} O_i$$

$$\Rightarrow \frac{\partial H_t}{\partial net_j} = \boxed{\sum_{k\in Downstream(j)} \frac{\partial H_t}{\partial net_k} \cdot \left(W_{kj} O_j - O_j \sum_i W_{ki} \cdot O_i\right)}$$

[例]-α-β


定下这个，才能有……>5
β>α 不剪
(α剪，因为根据父辈已记录的α)
(! β只能与α比) 此时根α未定
! 易漏 (根α已定)
! 易漏 β=2 < 根α=5


(与根比，3≤3，α剪)
(β剪，因15>β=3)
(<α=3)
(与根比)

[例1] 线性可分SVM，$x_1=(0,0)$ 正，$x_2=(2,0)$，$x_3=(0,2)$ 负，求一，对 $x_5=(0,-1)$ 分

$\min\limits_{\alpha} \frac{1}{2}(4\alpha_2^2+4\alpha_3^2)-(\alpha_1+\alpha_2+\alpha_3)$，s.t. $\alpha_1-\alpha_2-\alpha_3=0$，代入得

$s(\alpha_2,\alpha_3)=2\alpha_2^2+2\alpha_3^2-2\alpha_2-2\alpha_3$，

令偏导为0得 $\begin{cases}4\alpha_2-2=0\\4\alpha_3-2=0\end{cases} \Rightarrow \begin{cases}\alpha_2=\frac{1}{2}\\ \alpha_3=\frac{1}{2}\end{cases} \Rightarrow \alpha_1=1$. 均正 √ 且支持vec

$\Rightarrow$ 平面：$w^*=\alpha_1 y_1 x_1+\alpha_2 y_2 x_2+\alpha_3 y_3 x_3=-\frac{1}{2}\times(2,2)=(-1,-1)$

$b^*=y_1-(\alpha_1^* y_1(x_1\cdot x_1)+\alpha_2 y_2(x_2\cdot x_1)+\alpha_3 y_3(x_3\cdot x_1))=0$.

$-x_1-x_2=0$

$f(x)=\mathrm{sign}(-x_1-x_2) \Rightarrow f(x_5)=\mathrm{sign}(1)$ 为正例

[例2] 非线性SVM，$x_1=0$ 负，$x_2=1$，$x_3=-1$ 正，$K(x_i,x_j)=(x_i * x_j+1)^2$

$\cancel{\min\limits_{\alpha}\frac{1}{2}}$ ① 先算"内积"：$K(x_1,x_1)=1$，$K(x_1,x_2)=1$，$K(x_1,x_3)=1$，

$K(x_2,x_3)=0$，$K(x_2,x_2)=4$，$K(x_3,x_3)=4$

② $\min\limits_{\alpha}\frac{1}{2}\sum_{i=1}^{3}\sum_{j=1}^{3}\alpha_i\alpha_j y_i y_j K(x_i,x_j)-\sum_{i=1}^{3}\alpha_i$，s.t. $\sum_{i=1}^{3}\alpha_i y_i=0$，$\alpha_i\geq 0$

即：

$\min\limits_{\alpha}\frac{1}{2}(\alpha_1^2+4\alpha_2^2+4\alpha_3^2-2\alpha_1\alpha_2-2\alpha_1\alpha_3)-\alpha_1-\alpha_2-\alpha_3$，代入 $-\alpha_1+\alpha_2+\alpha_3=0$

$\Rightarrow s(\alpha_2,\alpha_3)=\frac{1}{2}((\alpha_2+\alpha_3)^2+4\alpha_2^2+4\alpha_3^2-2(\alpha_2+\alpha_3)^2)-2\alpha_2-2\alpha_3$

$=\frac{1}{2}(4\alpha_2^2+4\alpha_3^2-\alpha_2^2-2\alpha_2\alpha_3-\alpha_3^2)-2\alpha_2-2\alpha_3$

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

$\mathrm{grad}\,s=0 \Rightarrow \begin{cases}3\alpha_2-\alpha_3-2=0\\3\alpha_3-\alpha_2-2=0\end{cases} \Rightarrow \begin{cases}\alpha_2=1\\ \alpha_3=1\end{cases} \Rightarrow \alpha_1=2$ √

$\Rightarrow \cancel{w^*=0+1-1=0}$，

$\cancel{b^*=-1-(-2\times1+1\times1+1\times1)=-1}$.

超平面：

w^* 没有了，前一项是 $\sum_{i=1}^{N}\alpha_i^* y_i K(x_i,x)=\underbrace{\alpha_1 y_1}_{-2}(x_1\cdot x+1)^2+\underbrace{\alpha_2 y_2}_{1}(x_2\cdot x+1)^2+\underbrace{\alpha_3 y_3}_{1}(x_3\cdot x+1)^2$

$b^*=y_1-(\alpha_1 y_1 K(x_1,x_1)+\alpha_2 y_2 K(x_1,x_2)+\alpha_3 y_3 K(x_1,x_3))$

$=-1-(-2\times1+1\times1+1\times1)=-1$.

$\Rightarrow f(x)=\mathrm{sign}(-2(\cdots)^2+(\ \ )^2+(\ \ )^2-1)$