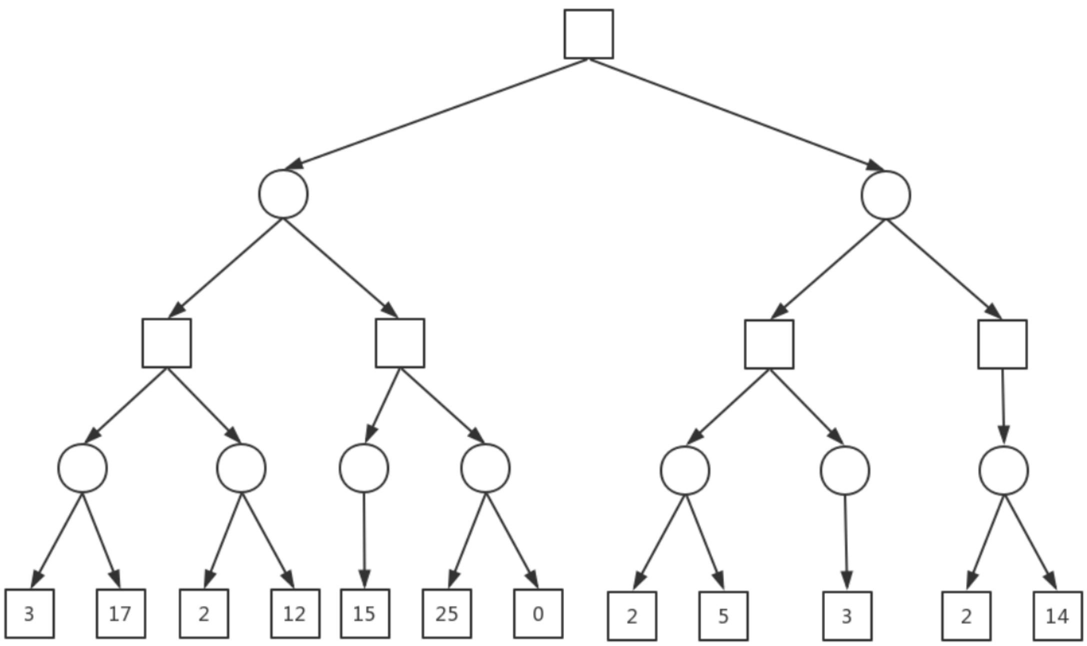
搜索

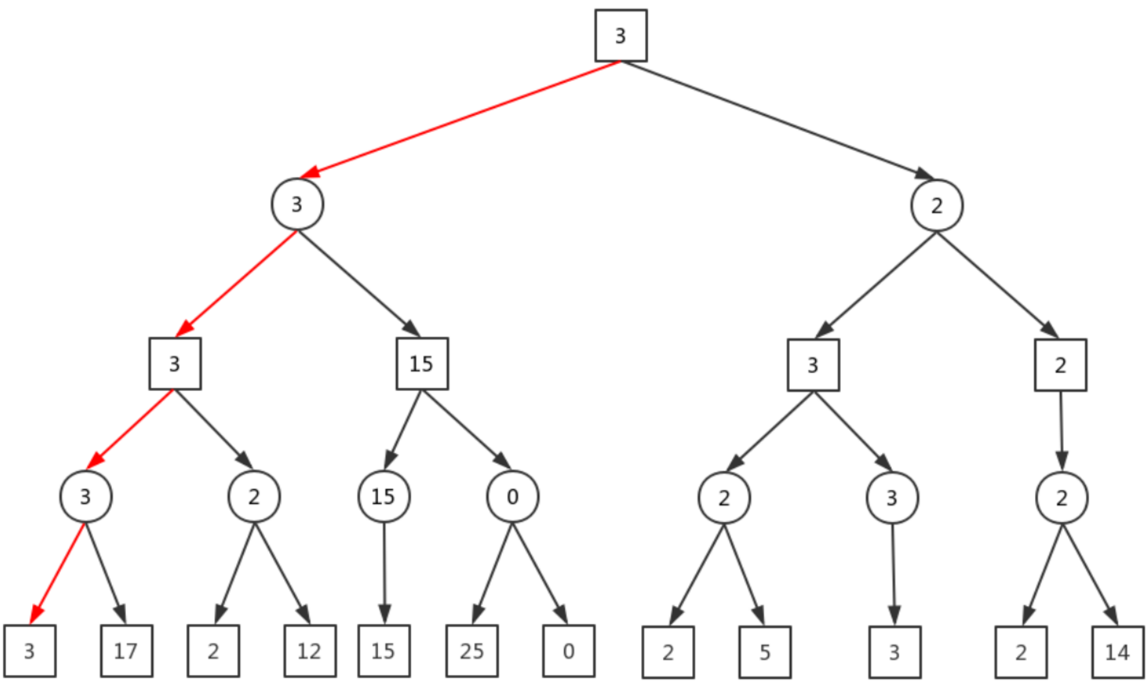
MinMax搜索

我们用**正方形来代表先手（选择估价最大的局面），圆形来代表后手（选择估价最小的局面）。只有叶子节点才可以直接计算估价值。**

则我们**假设棋局的博弈树**如下(**往后看4步**)，那么先手应该如何选择呢：



所以，如果先后手都进行最有决策的话，棋局的走向则下图所示（红线）：



Alpha-Beta剪枝

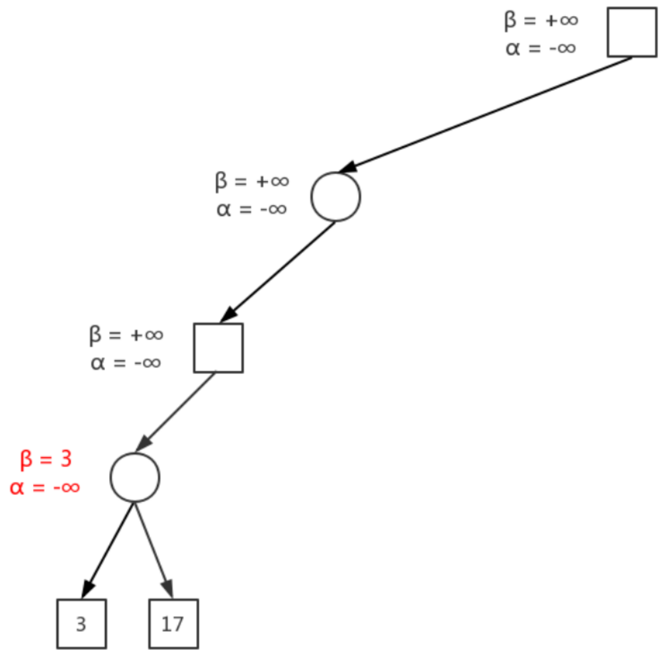
Alpha-beta($\alpha - \beta$)剪枝的名称来自计算过程中传递的两个边界，这些边界基于已经看到的搜索树部分来限制可能的解决方案集。 其中，Alpha(α)表示**目前**所有可能解中的最大下界，Beta(β)表示**目前**所有可能解中的最小上界。

因此，如果搜索树上的一个节点被考虑作为最优解的路上的节点（或者说是这个节点被认为是有必要进行搜索的节点），那么它一定满足以下条件（N是当前节点的估价值）：

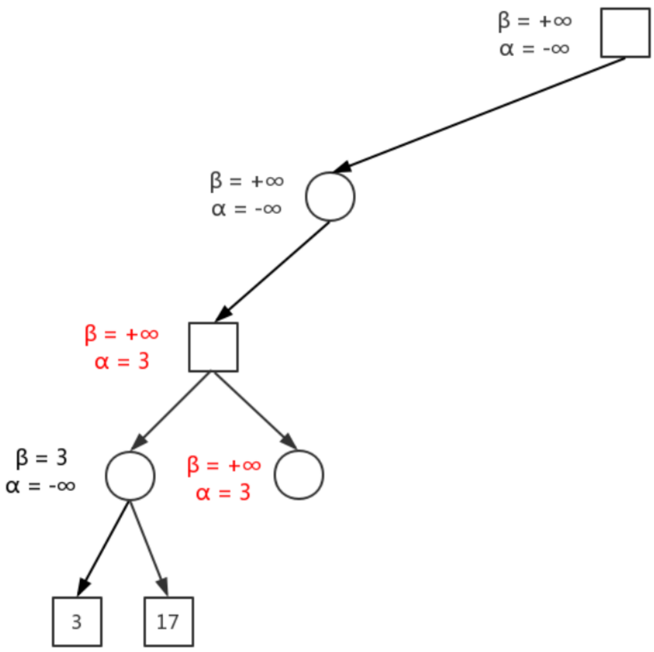
$$\alpha \leq \mathrm{N} \leq \beta$$

在我们进行求解的过程中，α和β会逐渐逼近。如果对于某一个节点，出现了$\alpha > \beta$的情况，那么，说明这个点一定不会产生最优解了，所以，我们就不再对其进行扩展（也就是不再生成子节点），这样就完成了对博弈树的剪枝。

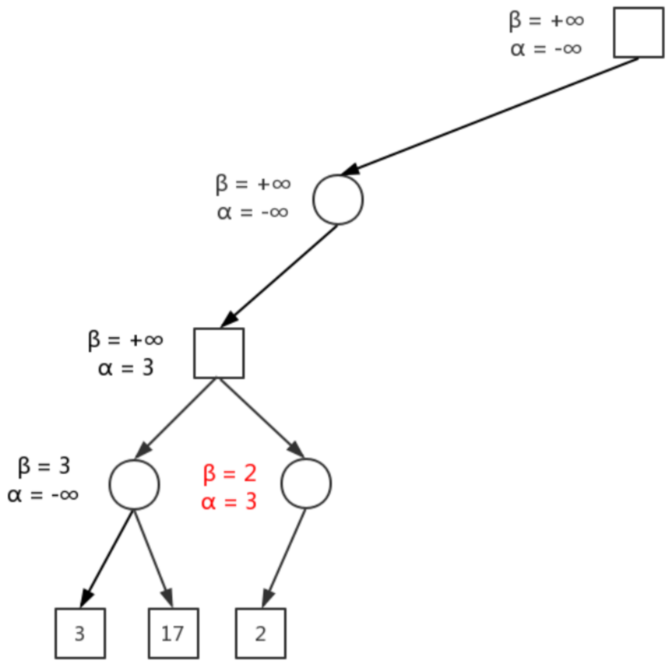
当我们搜索到了第一个叶子节点的时候，我们发现它的权值是3，并且它的父节点是Min节点，又因为父节点的最小上界$\beta > 3$，所以我们更新父节点，令$\beta = 3$。（因为父节点要取最小值，这个最小值不会比3更大，所以我们更新其上界）然后继续向下搜索。



因为17比3大，所以这个节点我们可以无视掉（~~没啥用啊~~）。我们已经搜索完了当前这个Min节点的所有孩子，所以我们返回它的节点值给它的父节点（Max节点），尝试更新父节点的α值。（因为这是一个Max节点，他的孩子的估价值和α、β值已经确定了，所以父节点取值范围的下界也需要被更新），此处更新父节点，令$\alpha = 3$。然后继续进行搜索，**注意新生成的子节点的** α、β **值继承自父节点。**

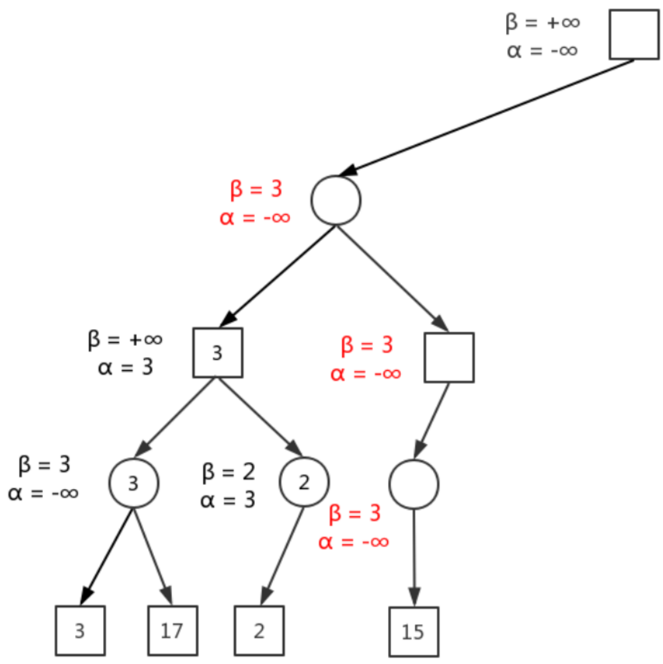


继续搜索，至叶子节点：



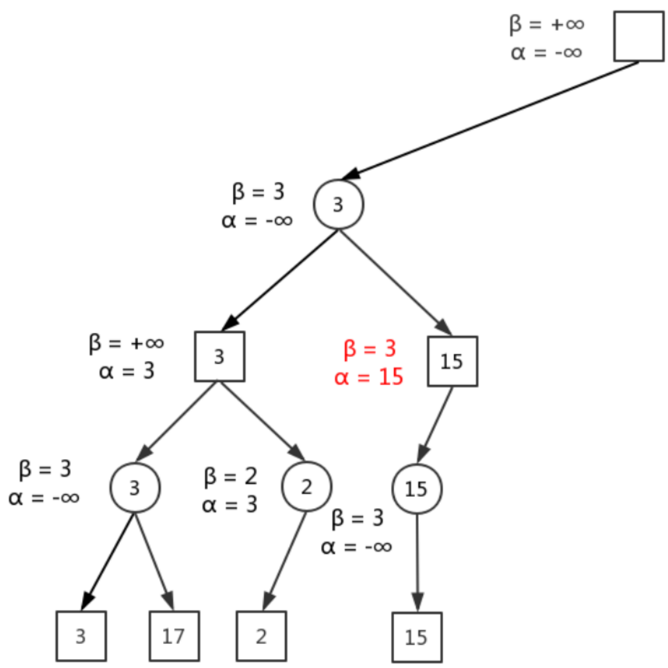
我们发现它的权值是2，并且它的父节点是Min节点，又因为父节点的最小上界$\beta > 2$，所以我们更新父节点，令$\beta = 2$。（因为父节点要取最小值，这个最小值不会比2更大，所以我们更新其上界）。然后此时我们发现父节点出现了$\alpha > \beta$的情况，说明最优解不可能从这个节点的子节点中产生，所以我们不再继续搜索它的其他子节点。（**这就是β剪枝**）并继续返回其节点值，尝试更新父节点。因为父节点的$\alpha = 3 > 2$，所以更新失败。

然后我们已经搜索完了当前这个Max节点的所有子节点，所以我们返回它的节点值，并尝试更新他的父节点的β 值。因为父节点的$\beta > 3$，所以我们令$\beta = 3 $。并继续向下搜索至叶子节点，**注意新生成的子节点的** α、β **值继承自父节点。**

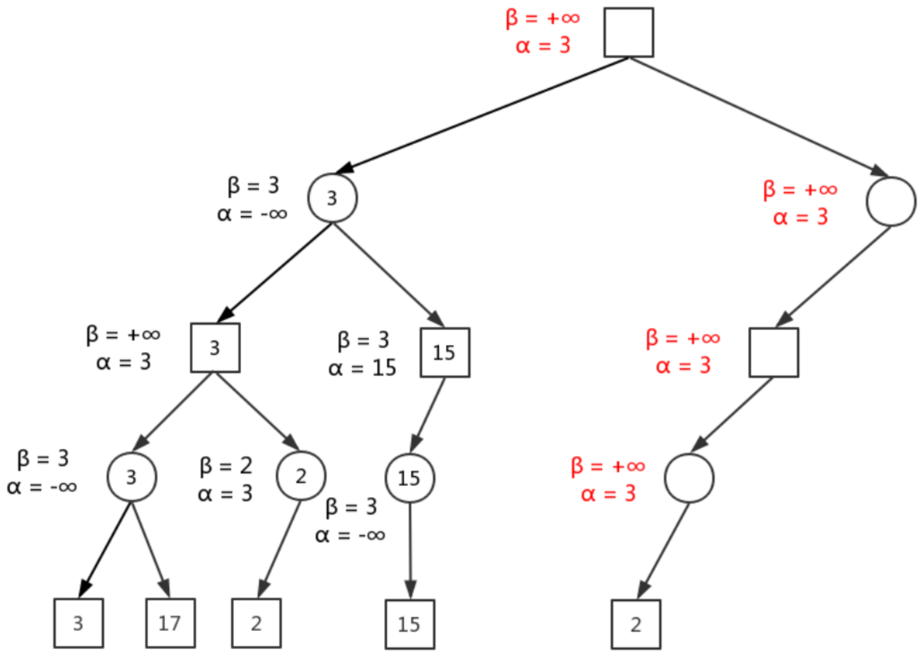


然后，我们发现15并不能更新当前节点的β值，所以令当前节点权值为15，并返回其权值，尝试更新其父节点(Max节点)的α值。因为其父节点的$\alpha < 15$，所以我们令$\alpha = 15 $。

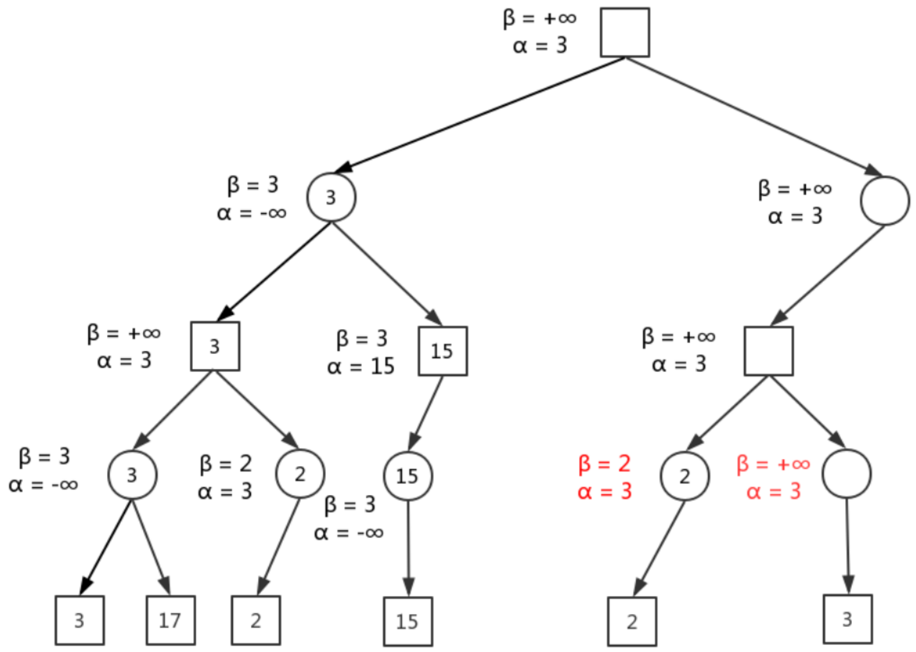
此时，该节点$\alpha = 15, \beta = 3$，$\alpha > \beta$，则说明其子节点并不包含最优解，不需要在进行搜索。所以返回其节点权值给父节点（Min节点），尝试对父节点的β值进行更新。父节点的$\beta < 15$，则不需要进行更新。同时可确定父节点的权值为3。



继续返回权值给父节点，尝试更新父节点的α，发现父节点$\alpha < 3$，则令$\alpha = 3$，并继续向下搜索直至叶子节点。**注意新生成的子节点的 α、β 值继承自父节点。**



从叶子节点返回权值给父节点（Min节点），并尝试更新其父节点的β值，因为父节点$\beta > 2$，所以，令$\beta = 2$，此时有$\alpha > \beta$，说明其子节点并不包含最优解，不需要再进行搜索。所以返回其节点权值给父节点（Max节点），尝试对父节点的α值进行更新。因为父节点$\alpha > 2$，无需进行更新，继续搜索其子节点至叶子节点。

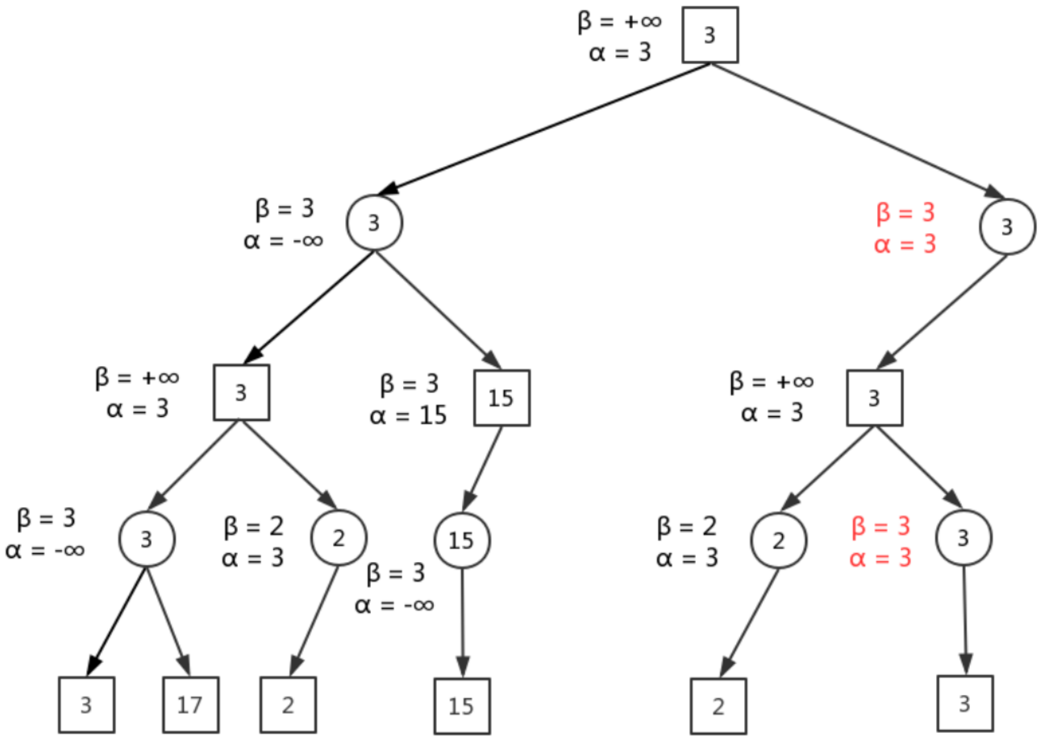


从叶子节点返回权值给父节点（Min节点），并尝试更新其父节点的β值，因为父节点$\beta > 3$，所以，令$\beta = 3$，同时确认父节点权值为3。

继续返回权值给父节点，并尝试更新其父节点的α值，因为父节点$\alpha = 3$，所以无需进行更新，同时确定该节点权值为3。

因为该节点的所有子节点全部搜索完毕，所以返回该点权值给父节点，并尝试更新其父节点的β值，因为父节点$\beta > 3$，所以，令$\beta = 3$，同时确认父节点权值为3。因为此时有$\alpha = \beta = 3$，所以无需再搜索其子节点，直接返回权值给根节点，并尝试更新根节点的α值，因为根节点$\alpha = 3$，所以无需进行更新。

根节点的所有子节点搜索完毕，则得出最优解为3。



推理

贝叶斯网络

FOIL（First Order Inductive Learner）算法思路：逐一添加前提约束谓词

表 2.7　添加前提约束谓词所得信息增益值计算过程

推理规则		推理规则涵盖的正例和反例数		FOIL 信息增益值
目标谓词	前提约束谓词	正例	反例	信息增益值
$Father(x, y) \leftarrow$	空集	$m_+ = 1$	$m_- = 4$	/
	$Mother(x, y)$	$\widehat{m_+} = 0$	$\widehat{m_-} = 2$	NA
	$Mother(x, z)$	$\widehat{m_+} = 0$	$\widehat{m_-} = 2$	NA
	$Mother(y, x)$	$\widehat{m_+} = 0$	$\widehat{m_-} = 1$	NA
	$Mother(y, z)$	$\widehat{m_+} = 0$	$\widehat{m_-} = 1$	NA
	$Mother(z, x)$	$\widehat{m_+} = 0$	$\widehat{m_-} = 1$	NA
	$Mother(z, y)$	$\widehat{m_+} = 1$	$\widehat{m_-} = 3$	0.32
	$Sibling(x, y)$	$\widehat{m_+} = 0$	$\widehat{m} = 1$	NA
	$Sibling(x, z)$	$\widehat{m_+} = 0$	$\widehat{m_-} = 1$	NA
	$Sibling(y, x)$	$\widehat{m_+} = 0$	$\widehat{m_-} = 0$	NA
$Father(x, y) \leftarrow$	$Sibling(y, z)$	$\widehat{m_+} = 0$	$\widehat{m_-} = 1$	NA
	$Sibling(z, x)$	$\widehat{m_+} = 0$	$\widehat{m_-} = 0$	NA
	$Sibling(z, y)$	$\widehat{m_+} = 1$	$\widehat{m_-} = 2$	0.74
	$Couple(x, y)$	$\widehat{m_+} = 0$	$\widehat{m_-} = 1$	NA
	$Couple(x, z)$	$\widehat{m_+} = 1$	$\widehat{m_-} = 1$	1.32
	$Couple(y, x)$	$\widehat{m_+} = 0$	$\widehat{m_-} = 0$	NA
	$Couple(y, z)$	$\widehat{m_+} = 0$	$\widehat{m_-} = 0$	NA
	$Couple(z, x)$	$\widehat{m_+} = 0$	$\widehat{m_-} = 2$	NA
	$Couple(z, y)$	$\widehat{m_+} = 0$	$\widehat{m_-} = 1$	NA

- 给定目标谓词$Father(x, y)$，此时推理规则只有目标谓词，因此推理规则所覆盖正例和反例样本数分别是训练样本中正例和反例的数目，即1和4，因此$m_+ = 1$和$m_- = 4$。
- 如果将$Mother(x, y)$作为前提约束谓词，加入推理规则，可得到$Mother(x, y) \rightarrow Father(x, y)$。在表中所提供背景知识样例集合中，$Mother(x, y)$有两个实例，即$Mother$(James, Ann)和$Mother$(James, Mike)。对于$Mother$(James, Ann)这一实例，$x$ =James和y =Ann，将x和y代入$Father(x, y)$得到$Father$(James, Ann)，从表2.4.2所提供的训练样本集合知道$Father$(James, Ann)是一个反例；对于$Mother$(James, Mike)这一实例，x =James，y =Mike，将x和y代入$Father(x, y)$得到$Father$(James, Mike)，从表中所提供的训练样本集合可知$Father$(James, Mike)是一个反例。因此推理规则$Mother(x, y) \rightarrow Father(x, y)$所覆盖的正例和反例数量分别为0和2，即$\widehat{m_+} = 0$，$\widehat{m_-} = 2$。由于$\widehat{m_+} = 0$，代入$FOIL_Gain$公式时会出现负无穷的情况，此时$FOIL_Gain$记为NA（Not Available）。
- 类似地，可将背景知识样例中出现的所有谓词（$Mother(\cdot,\cdot)$、$Sibling(\cdot,\cdot)$和$Couple(\cdot,\cdot)$）依次作为前提约束加入推理规则，如上计算得到推理规则的$FOIL_Gain$取值。

贝叶斯网络： 计算P(下雨,洒水车,路湿|多云)的概率

$P(\text{下雨},\text{洒水车},\text{路湿}|\text{多云}) = \frac{P(\text{下雨},\text{洒水车},\text{路湿},\text{多云})}{P(\text{多云})} =$

$\frac{P(\text{多云})P(\text{洒水车}|\text{多云})P(\text{下雨}|\text{多云})P(\text{路湿}|\text{洒水车},\text{下雨})}{P(\text{多云})} = \frac{0.4\times0.1\times0.8\times0.99}{0.4} = 0.0792$。计算P(路湿|不多云,

下雨)的概率，可以通过$P(\text{路湿}|\text{不多云},\text{下雨}) = \frac{P(\text{路湿},\text{不多云},\text{下雨})}{P(\text{不多云},\text{下雨})} =$

$\frac{P(\text{路湿},\text{不多云},\text{下雨},\text{洒水车})+P(\text{路湿},\text{不多云},\text{下雨},\text{无洒水车})}{P(\text{下雨}|\text{不多云})P(\text{不多云})} =$

$\frac{P(\text{不多云})P(\text{洒水车}|\text{不多云})P(\text{下雨}|\text{不多云})P(\text{路湿}|\text{洒水车},\text{下雨})+P(\text{不多云})P(\text{无洒水车}|\text{不多云})P(\text{下雨}|\text{不多云})P(\text{路湿}|\text{无洒水车},\text{下雨})}{P(\text{下雨}|\text{不多云})P(\text{不多云})} =$

$\frac{0.6\times0.5\times0.2\times0.99+0.6\times0.5\times0.2\times0.9}{0.2\times0.6} = 0.945$。

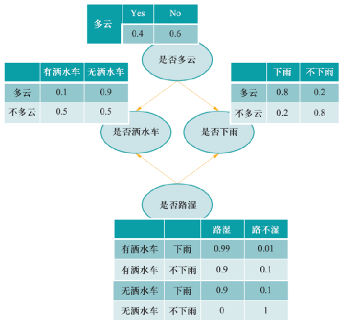


图 2.3 贝叶斯网络示例

无监督学习

LDA计算

计算步骤【二分类问题】：

1. 计算类间散度矩阵 S_b

$$S_b = (\mu_0 - \mu_1)(\mu_0 - \mu_1)^T$$

   其中 μ_0 为标签为 0 的特征平均值，其值个数等于特征个数，μ_1 同理

2. 计算类内散度矩阵 S_ω

$$S_\omega = \Sigma_0 + \Sigma_1$$

   其中 Σ_0 是标签为 0 的样本协方差矩阵，若有 n 个特征，则其大小为 $n \times n$

$$\Sigma_0 = \begin{bmatrix} Cov(f_1, f_1) & Cov(f_1, f_2) & \cdots & Cov(f_1, f_n) \\ Cov(f_2, f_1) & Cov(f_2, f_2) & \cdots & Cov(f_2, f_n) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(f_n, f_1) & Cov(f_n, f_2) & \cdots & Cov(f_n, f_n) \end{bmatrix}$$

   其中 $Cov(x_1, x_2) = \frac{1}{n-1} \sum_{i=1}^{n} (x_1^i - \overline{x_1})(x_2^i - \overline{x_2})$，$n$ 为样本个数

3. 计算矩阵 $S_\omega^{-1} S_b$

   逆矩阵可以使用初等变换辅助求解，变换规则如下：

   - 对调矩阵两行（列）
   - 矩阵某行（列）乘以非零常数 k
   - 矩阵某行（列）倍数加到另一行

4. 对 $S_\omega^{-1}S_b$ 矩阵求特征值和特征向量，选择特征值最大的特征向量作为 ω【需要归一化】

   - 特征值可通过 $|\lambda E - A| = 0$ 求得
   - 将特征值带入方程 $(\lambda_0 E - A)X = 0$，非零解即为特征值 λ_0 对应得特征向量

   > 具体可参考考研数学线性代数部分，不再赘述

5. 计算得到投影后的数据点 $Y = X\omega$

   将样本值 X 代入，得到的结果即投影后对应的位置

题目

假设有如下 10 个样本，样本有 2 个特征，前 5 项为负类，后 5 项为正类

$$\begin{aligned} D &= \{X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8, X_9, X_{10}\} \\ &= \{(4,2)^T, (2,4)^T, (2,3)^T, (3,6)^T, (4,4)^T, (9,10)^T, (6,8)^T, (9,5)^T, (8,7)^T, (10,8)^T\} \end{aligned}$$

计算当前样本的类间散度矩阵 S_b 和类内散度矩阵 S_ω

题目所给样本可组成如下矩阵👇

样本	特征值1	特征值2	分类
X_1	4	2	0
X_2	2	4	0
X_3	2	3	0
X_4	3	6	0
X_5	4	4	0
X_6	9	10	1
X_7	6	8	1
X_8	9	5	1
X_9	8	7	1
X_{10}	10	8	1

$$
\begin{aligned}
\mu_0 &= [(\frac{4+2+2+3+4}{5}),(\frac{2+4+3+6+4}{5})]^T = [3, 3.8]^T \\
\therefore \mu_1 &= [8.4, 7.6]^T \\
\therefore S_b &= (\mu_0-\mu_1)(\mu_0-\mu_1)^T \\
&= \begin{bmatrix} -5.4 \\ -4.2 \end{bmatrix} \begin{bmatrix} -5.4 & -4.2 \end{bmatrix} \\
&= \begin{bmatrix} 29.16 & 22.68 \\ 22.68 & 17.64 \end{bmatrix} \\
S_\omega &= \Sigma_0 + \Sigma_1 \\
\Sigma_0 &= \begin{bmatrix} Cov(f_1,f_1) & Cov(f_1,f_2) \\ Cov(f_2,f_1) & Cov(f_2,f_2) \end{bmatrix}
\end{aligned}
$$

以求 $Cov(f_1,f_2)$ 为例，如下👇，由于是对标签为 0 的样本计算协方差，$\therefore \overline{f_1}=3$，$\overline{f_2}=3.8$

$$
\begin{aligned}
Cov(f_1,f_2) &= \frac{1}{5-1}\sum_{i=1}^{5}(f_1^i-\overline{f_1})(f_2^i-\overline{f_2}) \\
&= \frac{1}{4}[(4-3)(2-3.8)+(2-3)(4-3.8)+(2-3)(3-3.8)+(3-3)(6-3.8) \\
&\quad +(4-3)(4-3.8)] \\
&= \frac{1}{4}\times(-1.8-1.8+0.8+1.8) \\
&= -0.25
\end{aligned}
$$

其他数值都可根据类似方法得到

$$
\begin{aligned}
\therefore \Sigma_0 &= \begin{bmatrix} 1 & -0.25 \\ -0.25 & 2.2 \end{bmatrix} \\
\Sigma_1 &= \begin{bmatrix} 2.3 & -0.05 \\ -0.05 & 3.3 \end{bmatrix} \\
\therefore S_\omega = \Sigma_0 + \Sigma_1 &= \begin{bmatrix} 3.3 & -0.3 \\ -0.3 & 5.5 \end{bmatrix}
\end{aligned}
$$

PCA（主成分分析）

例一

给定X的协差阵，对其进行主成分分析，

$$\Sigma=\begin{bmatrix}2 & 0 & 0\\ 0 & 3/2 & 1/2\\ 0 & 1/2 & 3/2\end{bmatrix}$$

（1）求出每个主成分的贡献率;
（2）求出每个原始变量的信息提取率;

解：对于主成分分析的题，一般来说，题目给定一个协方差阵，不管怎样先求出特征值和特征向量。

`Step1` 计算特征根

解$|\Sigma-\lambda I|$=0，得：λ_1=2，λ_2=2，λ_3=1 （$\lambda_1 \geq \lambda_2 \geq \lambda_3$）
如果解出来不确定或者解不出来的话可以通过 特征值的和等于协方差阵对角线元素的和 以及 特征值的积等于协方差阵对应的行列式 来进行验证

`Step2` 求特征向量，这里一定不能忘记要化成单位特征向量

$$U_1=\begin{bmatrix}1\\0\\0\end{bmatrix}\quad U_2=\begin{bmatrix}0\\ \frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}}\end{bmatrix}\quad U_3=\begin{bmatrix}0\\ -\frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}}\end{bmatrix}$$

`Step3` 计算贡献率

第一个主成分的贡献率为：$\lambda_1/(\lambda_1+\lambda_2+\lambda_3)$=2/5=40%
第二个主成分的贡献率为：$\lambda_2/(\lambda_1+\lambda_2+\lambda_3)$=2/5=40%
第三个主成分的贡献率为：$\lambda_3/(\lambda_1+\lambda_2+\lambda_3)$=1/5=20%
（注意这里算的不是累积贡献率）所以我们取前两个主成分就可以了

`Step4` 求信息提取率

分别计算x_i与F_1、F_2的相关系数的平方，信息提取率为两者之和，
计算相关系数的公式如下

$$\rho(x_i,F_j)=\frac{u_{ij}\lambda_j}{\sigma_i\sqrt{\lambda_j}}=\frac{u_{ij}\sqrt{\lambda_j}}{\sigma_i}$$

x_i	x_i与F_1相关系数的平方	x_i与F_2相关系数的平方	信息提取率
1	1	0	1
2	0	2/3	0.67
3	0	2/3	0.67

例二

设x_1，x_2，x_3的**协方差矩阵**如下，试求主成分分析，并求出每个主成分的**贡献率**及每个原始变量的**信息提取率**。

$$\Sigma=\begin{bmatrix}1 & -2 & 0\\ -2 & 5 & 0\\ 0 & 0 & 2\end{bmatrix}$$

解:

`Step1` 计算特征根

解$|\Sigma-\lambda I|$=0，得：λ_1=5.83，λ_2=2，λ_3=0.17 （λ_1≥λ_2≥λ_3）

`Step2` 求特征向量

$$U_1=\begin{bmatrix}0.383\\ -0.924\\ 0.000\end{bmatrix}\ U_2=\begin{bmatrix}0\\ 0\\ 1\end{bmatrix}\ U_3=\begin{bmatrix}0.924\\ 0.383\\ 0.000\end{bmatrix}$$

`Step3` 计算贡献率
第一个主成分的贡献率为：λ_1/(λ_1+λ_2+λ_3)=5.83/8=72.875%
第二个主成分的贡献率为：λ_2/(λ_1+λ_2+λ_3)=2/8=25%
第三个主成分的贡献率为：λ_3/(λ_1+λ_2+λ_3)=0.17/8=2.125%

`Step4` 求信息提取率
虽然第一个主成分的贡献率不小，但在本题中第一主成分不含第三个原始变量的信息，因此应该取两个主成分
所以分别计算x_i与F_1 F_2的相关系数的平方，信息提取率为两者之和

x_i	x_i与F_1相关系数的平方	x_i与F_2相关系数的平方	信息提取率
1	0.855	0	0.855
2	0.996	0	0.996
3	0	1	1