※卡方检测：

卡方检验就是统计样本的实际观测值与理论推断值之间的偏离程度，实际观测值与理论推断值之间的偏离程度就决定卡方值的大小，如果卡方值越大，二者偏差程度越大；反之，二者偏差越小；若两个值完全相等时，卡方值就为 0，表明理论值完全符合。

Ex：我们有一堆新闻标题，需要判断标题中包含某个词（比如吴亦凡）是否与该条新闻的类别归属（比如娱乐）是否有关，我们只需要简单统计就可以获得这样的一个四格表：

标题是否包含吴亦凡确实对新闻是否属于娱乐有统计上的差别，包含吴亦凡的新闻属于

组别	属于娱乐	不属于娱乐	合计
不包含吴亦凡	19	24	43
包含吴亦凡	34	10	44
合计	53	34	87

娱乐的比例更高，但我们还无法排除这个差别是否由于抽样误差导致。那么首先假设标题是否包含吴亦凡与新闻是否属于娱乐是独立无关的，随机抽取一条新闻标题，属于娱乐类别的概率是：(19 + 34) / (19 + 34 + 24 +10) = 60.9%

根据无关性假设生成新的理论值四格表：

组别	属于娱乐	不属于娱乐	合计
不包含吴亦凡	43 * 0.609 = 26.2	43 * 0.391 = 16.8	43
包含吴亦凡	44 * 0.609 = 26.8	44 * 0.391 = 17.2	44

x^2的计算公式为：$x^2=\sum\frac{(A-T)^2}{T}$ 也称拟合度公式 $\sum_{i=1}^{k}\frac{(f_i-np_i)^2}{np_i}$

A 为实际值，也就是第一个四格表里的 4 个数据。T 为理论值，也就是理论值四格表里的 4 个数据。对上述场景可计算x^2值为 10.01。

将x^2与临界分布表对比，临界分布表一般自己随便写或者题目给。

※相关系数（皮尔逊系数）：

方差： $\sigma^2=\frac{\sum(X-\mu)^2}{N}$ σ^2 为总体方差，X 为变量，μ 为总体均值，N 为总体例数。

$S^2=\frac{\sum(X-\overline{X})^2}{n-1}$ S^2 为样本方差，X 为变量，$\overline{X}$ 为样本均值，n 为样本例数。

期望值分别为*E*[*X*]与*E*[*Y*]的两个实随机变量*X*与*Y*之间的**协方差***Cov*(*X*, *Y*)定义为：

协方差：

$$\begin{aligned} Cov(X,Y) &= E[(X-E[X])(Y-E[Y])] \\ &= E[XY]-2E[Y]E[X]+E[X]E[Y] \\ &= E[XY]-E[X]E[Y] \end{aligned}$$

如果两组数据$X:\{X_1,X_2,\cdots,X_n\}$和$Y:\{Y_1,Y_2,\cdots,Y_n\}$是总体数据（例如普查结果），

那么总体均值：$E(X)=\frac{\sum_{i=1}^{n}X_i}{n}$，$E(Y)=\frac{\sum_{i=1}^{n}Y_i}{n}$

总体协方差：$\mathrm{Cov}(X,Y)=\frac{\sum_{i=1}^{n}(X_i-E(X))(Y_i-E(Y))}{n}$

相关系数定义式：

$$r(X,Y)=\frac{Cov(X,Y)}{\sqrt{Var[X]Var[Y]}}$$

Cov(X,Y)为 X 与 Y 的协方差，Var[X]为 X 的方差，Var[Y]为 Y 的方差

※PCA：

原理：PCA(Principal Component Analysis)，即主成分分析方法，是一种使用最广泛的数据降维算法。PCA 的主要思想是将 n 维特征映射到 k 维上，这 k 维是全新的正交特征也被称为主成分，是在原有 n 维特征的基础上重新构造出来的 k 维特征。PCA 的工作就是从原始的空间中顺序地找一组相互正交的坐标轴，新的坐标轴的选择与数据本身是密切相关的。其中，第一个新坐标轴选择是原始数据中方差最大的方向，第二个新坐标轴选取是与第一个坐标轴正交的平面中使得方差最大的，第三个轴是与第 1,2 个轴正交的平面中方差最大的。依次类推，可以得到 n 个这样的坐标轴。

算法流程：

输入：n为样本集$D=(x^{(1)},x^{(2)},...,x^{(m)})$，设为 X，需要降维到 n'

输出：降维后的样本集 D'

(1) 先对所有的数据集进行中心化：$x^{(i)}=x^{(i)}-\frac{1}{m}\sum_{j=1}^{m}x^{(j)}$

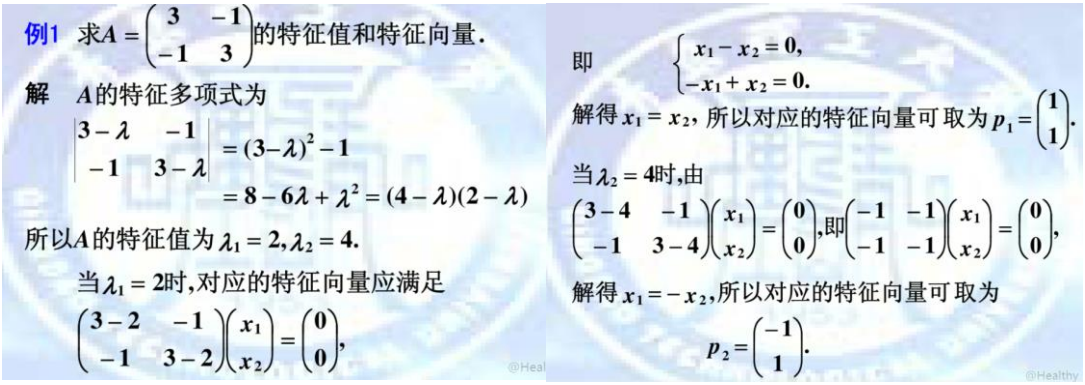
(2) 计算样本的协方差矩阵：XX^T

(3) 对协方差矩阵进行求特征值和特征向量

(4) 取出n'个最大的特征值对应的特征向量$(w_1,w_2,...,w_{n'})$，将所有的特征向量标准化，组成新的矩阵 w

(5) 输出矩阵：Y= WX 即为降 维到 n' 维后的数据

※特征值和特征向量求法：



例1 求$A=\begin{pmatrix}3&-1\\-1&3\end{pmatrix}$的特征值和特征向量.

解 A的特征多项式为

$$\begin{vmatrix}3-\lambda&-1\\-1&3-\lambda\end{vmatrix}=(3-\lambda)^2-1$$
$$=8-6\lambda+\lambda^2=(4-\lambda)(2-\lambda)$$

所以A的特征值为$\lambda_1=2,\lambda_2=4.$

当$\lambda_1=2$时,对应的特征向量应满足

$$\begin{pmatrix}3-2&-1\\-1&3-2\end{pmatrix}\begin{pmatrix}x_1\\x_2\end{pmatrix}=\begin{pmatrix}0\\0\end{pmatrix},$$

即 $\begin{cases}x_1-x_2=0,\\-x_1+x_2=0.\end{cases}$

解得$x_1=x_2$，所以对应的特征向量可取为$p_1=\begin{pmatrix}1\\1\end{pmatrix}.$

当$\lambda_2=4$时,由

$$\begin{pmatrix}3-4&-1\\-1&3-4\end{pmatrix}\begin{pmatrix}x_1\\x_2\end{pmatrix}=\begin{pmatrix}0\\0\end{pmatrix},\text{即}\begin{pmatrix}-1&-1\\-1&-1\end{pmatrix}\begin{pmatrix}x_1\\x_2\end{pmatrix}=\begin{pmatrix}0\\0\end{pmatrix},$$

解得$x_1=-x_2$,所以对应的特征向量可取为

$$p_2=\begin{pmatrix}-1\\1\end{pmatrix}.$$

※协方差矩阵：

(1) 方差的计算公式是针对一维特征，即针对同一特征不同样本的取值来进行计算得到；而协方差则必须要求至少满足二维特征；方差是协方差的特殊情况。

(2) 方差和协方差的除数是n-1,这是为了得到方差和协方差的无偏估计。

协方差为正时，说明X和Y是正相关关系；协方差为负时，说明X和Y是负相关关系；协方差为0时，说明X和Y是相互独立。Cov(X,X)就是X的方差。当样本是n维数据时，它们的协方差实际上是协方差矩阵(对称方阵)。例如，对于3维数据(x,y,z)，计算它的协方差就是：

$$Cov(X,Y,Z)=\begin{bmatrix}Cov(x,x)&Cov(x,y)&Cov(x,z)\\Cov(y,x)&Cov(y,y)&Cov(y,z)\\Cov(z,x)&Cov(z,y)&Cov(z,z)\end{bmatrix}$$

Ex：现在假设有一组数据如下：

	x	y
	2.5	2.4
	0.5	0.7
	2.2	2.9
	1.9	2.2
Data =	3.1	3.0
	2.3	2.7
	2	1.6
	1	1.1
	1.5	1.6
	1.1	0.9

行代表了样例，列代表特征，这里有 10 个样例，每个样例两个特征。可以这样认为，有 10 篇文档，x 是 10 篇文档中“learn”出现的 TF-IDF，y 是 10 篇文档中“study”出现的 TF-IDF。

分别求 x 和 y 的平均值，然后对于所有的样例，都减去对应的均值。这里 x 的均值是 1.81，y 的均值是 1.91，那么一个样例减去均值后即为（0.69,0.49），**得到**

※求特征协方差矩阵

	x	y
	0.69	0.49
	-1.31	-1.21
	0.39	0.99
	0.09	0.29
DataAdjust =	1.29	1.09
	0.49	0.79
	0.19	-0.31
	-0.81	-0.81
	-0.31	-0.31
	-0.71	-1.01

$$cov=\begin{pmatrix}.616555556&.615444444\\.615444444&.716555556\end{pmatrix}$$

对角线上分别是 x 和 y 的方差，非对角线上是协方差。协方差是衡量两个变量同时变化的变化程度。协方差大于 0 表示 x 和 y 若一个增，另一个也增；小于 0 表示一个增，一个减。如果 x 和 y 是统计独立的，那么二者之间的协方差就是 0；但是协方差是 0，并不能说明 x 和 y 是独立的。协方差绝对值越大，两者对彼此的影响越大，反之越小。协方差是没有单位的量，因此，如果同样的两个变量所采用的量纲发生变化，它们的协方差也会产生树枝上的变化。

求协方差的特征值和特征向量，得到

$$eigenvalues = \begin{pmatrix} .0490833989 \\ 1.28402771 \end{pmatrix}$$

$$eigenvectors = \begin{pmatrix} -.735178656 & -.677873399 \\ .677873399 & -.735178656 \end{pmatrix}$$

上面是两个特征值，下面是对应的特征向量，特征值 0.0490833989 对应特征向量为，这里的特征向量都归一化为单位向量。

将特征值按照从大到小的顺序排序，选择其中最大的 k 个，然后将其对应的 k 个特征向量分别作为列向量组成特征向量矩阵。

这里特征值只有两个，我们选择其中最大的那个，这里是 1.28402771，对应的特征向量是(-0.677873399, -0.735178656)T

将样本点投影到选取的特征向量上。假设样例数为 m，特征数为 n，减去均值后的样本矩阵为 DataAdjust(m*n), 协方差矩阵是 n*n, 选取的 k 个特征向量组成的矩阵为 EigenVectors(n*k)。那么投影后的数据 FinalData 为

FinalData(10*1) = DataAdjust(10*2矩阵) x 特征向量(-0.677873399, -0.735178656)T

得到的结果是：

Transformed Data (Single rigenvector)
x
-0.827970186
1.77758033
-0.992197494
-0.274210416
-1.67580142
-0.912949103
0.991094375
1.14457216
0.438046137
1.22382056

※SVD 相关：

特征值分解和奇异值分解在机器学习中都是很常见的矩阵分解算法。两者有着很紧密的关

系，特征值分解和奇异值分解的目的都是一样，就是提取出一个矩阵最重要的特征.

※特征值分解：

特征值、特征向量

如果一个向量 v 是矩阵 A 的特征向量，将一定可以表示成下面的形式：

$$Av = \lambda v$$

其中，λ是特征向量 v 对应的特征值，一个矩阵的一组特征向量是一组正交向量。

一个矩阵其实就是一个线性变换，因为一个矩阵乘以一个向量后得到的向量，其实就相当于将这个向量进行了线性变换。比如说下面的这个矩阵：

$$M = \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix}$$

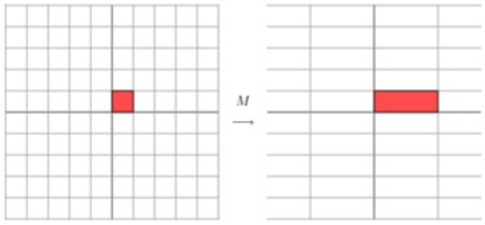


图2：矩阵M的线性变换

对于矩阵 A，有一组特征向量 v，将这组向量进行正交化单位化，就能得到一组正交单位向量。**特征值分解**，就是将矩阵 A 分解为如下式：　$A = Q\Sigma Q^{-1}$

其中，Q 是矩阵 A 的特征向量组成的矩阵，Σ 则是一个对角阵，对角线上的元素就是特征值。Σ矩阵是一个对角矩阵，里面的特征值是由大到小排列的，这些特征值所对应的特征向量就是描述这个矩阵变换方向

Ex：

这里我们用一个简单的方阵来说明特征值分解的步骤。我们的方阵 A 定义

$$A = \begin{pmatrix} -1 & 1 & 0 \\ -4 & 3 & 0 \\ 1 & 0 & 2 \end{pmatrix}$$

首先，由方阵 A 的特征方程，求出特征值。

$$|A - \lambda E| = \begin{vmatrix} -1-\lambda & 1 & 0 \\ -4 & 3-\lambda 0 & \\ 1 & 0 & 2-\lambda \end{vmatrix} = (2-\lambda)\begin{vmatrix} -1-\lambda & 1 \\ -4 & 3-\lambda \end{vmatrix} = (2-\lambda)(\lambda-1)^2 = 0$$

特征值为 $\lambda = 2, 1$ （重数是 2）。

然后，把每个特征值λ带入线性方程组$(A-\lambda E)x=0$，求出特征向量。

当λ=2时，解线性方程组 $(A-2E)x=0$。

$$(A-2E)=\begin{pmatrix}-3&1&0\\-4&1&0\\1&0&0\end{pmatrix}\rightarrow\begin{pmatrix}1&0&0\\0&1&0\\0&0&0\end{pmatrix}$$

解得 $x_1=0,\ x_2=0$。特征向量为：$p_1=\begin{pmatrix}0\\0\\1\end{pmatrix}$。

当λ=1时，解线性方程组 $(A-E)x=0$

$$(A-2E)=\begin{pmatrix}-2&1&0\\-4&2&0\\1&0&1\end{pmatrix}\rightarrow\begin{pmatrix}1&0&1\\0&1&2\\0&0&0\end{pmatrix}$$

$x_1+x_3=0,\ x_2+2x_3=0$。特征向量为：$p_2=\begin{pmatrix}-1\\-2\\1\end{pmatrix}$。

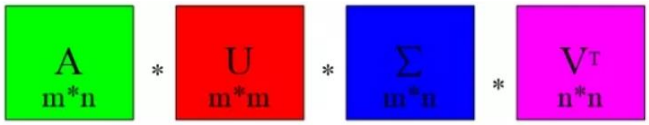
最后，方阵A的特征值分解为：

$$A=Q\Sigma Q^{-1}=\begin{pmatrix}0&-1&-1\\0&-2&-2\\1&1&1\end{pmatrix}\begin{pmatrix}2&0&0\\0&1&0\\0&0&1\end{pmatrix}\begin{pmatrix}0&-1&-1\\0&-2&-2\\1&1&1\end{pmatrix}^{-1}$$

奇异值分解：

奇异值分解是一个能适用于任意矩阵的一种分解的方法，对于任意矩阵 A 总是存在一个奇异值分解：$A=U\Sigma V^T$

假设 A 是一个 m*n 的矩阵，那么得到的 U 是一个 m*m 的方阵，U 里面的正交向量被称为左奇异向量。Σ是一个 m*n 的矩阵，Σ除了对角线其它元素都为 0，对角线上的元素称为奇异值。 V^T 是 v 的转置矩阵，是一个 n*n 的矩阵，它里面的正交向量被称为右奇异值向量。而且一般来讲，我们会将Σ上的值按从大到小的顺序排列。上面矩阵的维度变化可以参照图所示。



首先，我们用矩阵 A 的转置乘以 A，得到一个方阵，用这样的方阵进行特征分解，得到的特征值和特征向量满足下面的等式：$(A^TA)v_i=\lambda_i v_i$

这里的v_i就是我们要求的右奇异向量。

其次，我们将 A 和 A 的转置做矩阵的乘法，得到一个方阵，用这样的方阵进行特征分解，得到的特征和特征向量满足下面的等式：$(AA^T)u_i = \lambda_i u_i$

这里的 u_i 就是左奇异向量。

Ex：

这里我们用一个简单的矩阵来说明奇异值分解的步骤。我们的矩阵 A 定义为：

$$A = \begin{pmatrix} 0 & 1 \\ 1 & 1 \\ 1 & 0 \end{pmatrix}$$

首先，我们先求出 A^TA 和 AA^T：

$$A^TA = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$$

$$AA^T = \begin{pmatrix} 0 & 1 \\ 1 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 & 1 \\ 1 & 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 1 \end{pmatrix}$$

然后，求出 A^TA 和 AA^T 的特征值和特征向量：

A^TA 的特征值和特征向量：

$$\lambda_1 = 3;\ v_1 = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix};\ \lambda_2 = 1;\ v_2 = \begin{pmatrix} \frac{-1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix};$$

AA^T 的特征值和特征向量：

$$\lambda_1 = 3;\ u_1 = \begin{pmatrix} \frac{1}{\sqrt{6}} \\ \frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} \end{pmatrix};\ \lambda_2 = 1;\ u_2 = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ 0 \\ \frac{-1}{\sqrt{2}} \end{pmatrix};\ \lambda_3 = 0;\ u_3 = \begin{pmatrix} \frac{1}{\sqrt{3}} \\ \frac{-1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \end{pmatrix};$$

其次，我们利用 $Av_i = \sigma_i u_i,\ i = 1, 2$,求奇异值：

$$\begin{pmatrix} 0 & 1 \\ 1 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} = \sigma_1 \begin{pmatrix} \frac{1}{\sqrt{6}} \\ \frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} \end{pmatrix} \Rightarrow \sigma_1 = \sqrt{3}$$

$$\begin{pmatrix} 0 & 1 \\ 1 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} \frac{-1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} = \sigma_2 \begin{pmatrix} \frac{1}{\sqrt{2}} \\ 0 \\ \frac{-1}{\sqrt{2}} \end{pmatrix} \Rightarrow \sigma_2 = 1$$

当然，这一步也可以用 $\sigma_i = \sqrt{\lambda_i}$ 直接求出奇异值为 $\sqrt{3}$ 和1。

※十折交叉：

英文名叫做 10-fold cross-validation，用来测试算法准确性。是常用的测试方法。将数据集分成十分，轮流将其中 9 份作为训练数据，1 份作为测试数据，进行试验。每次试验都会得出相应的正确率（或差错率）。10 次的结果的正确率（或差错率）的平均值作为对算法精度的估计，一般还需要进行多次 10 折交叉验证（例如 10 次 10 折交叉验证），再求其均值，作为对算法准确性的估计。十折交叉验证之所以选择将数据集分为 10 份，是因为通过利用大量数据集、使用不同学习技术进行的大量试验，表明 10 折是获得最好误差

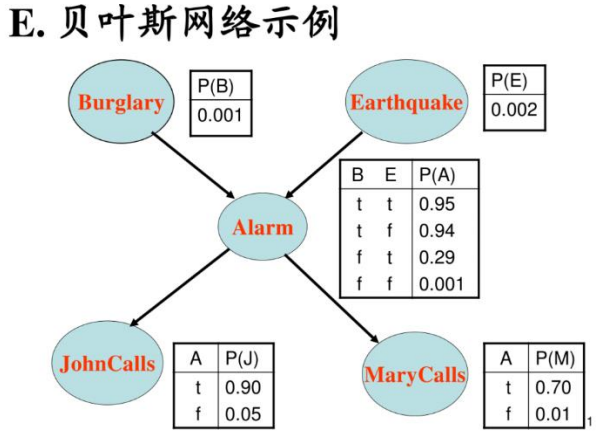
估计的恰当选择，而且也有一些理论根据可以证明这一点。但这并非最终诊断，争议仍然存在。而且似乎 5 折或者 20 折与 10 折所得出的结果也相差无几。

※二元分类：

二元分类又称“二向分类”。在包含两类事项的比较研究中，按两个标志所作的分类。
两元分类，这是一个专业预约的数学词语，也就是在它的方程中间包含有xy两种不同的元素。

※贝叶斯信念网络：之前我们讨论了朴素贝叶斯分类。朴素贝叶斯分类有一个限制条件，就是**特征属性必须有条件独立或基本独立**（实际上在现实应用中几乎不可能做到完全独立）。当这个条件成立时，朴素贝叶斯分类法的准确率是最高的，但是，现实中各个特征属性间往往并不条件独立，而是具有较强的相关性，这样就限制了朴素贝叶斯分类的能力。贝叶斯信念网络说明联合概率分布，它提供一种因果关系的图形，可以在其上进行学习。

※信念网络由两部分定义。第一部分是**有向无环图**，其每个结点代表一个随机变量，而每条弧代表一个概率依赖。如果一条弧由结点 Y 到 Z，则 Y 是 Z 的双亲或直接前驱，而 Z 是 Y 的后继。第二部分是每个属性一个**条件概率表**（CPT）。



示例：

说明：

Burglary 是盗贼闯入，其中 P(B)表示盗贼闯入的概率，为 0.001；

Earthquake 表示地震，其中 P(E)表示地震的概率，为 0.002；

Alarm 是报警器，其中 B E 对应的 t t 表示盗贼闯入且地震（t 表示 true，f 表示 false），其中 P(A)为 0.95，即盗贼闯入且地震情况下报警器响的概率为 0.95 其他以此类推；

JohnCalls 表示 John 打电话给你，意思是报警器响后 john 给你打电话，其中 A 就是报警器响条件下，P(J)给你打电话，其中 0.90 意思是 John 在你的报警器响后给你打电话的概率为 0.90，同理推倒 MaryCalls；

- 贝叶斯网络的语义

$$P(x_1,\ldots,x_n) = P(x_1|parent(x_1)) \ldots P(x_n|parent(x_n))$$

（重要公式，其中 parent(x1)表示 x1 的父节点，类似于 JohnCalls 的父节点 Alarm。同时各个 P(xn|…)之间相乘）

- 试计算：报警器响了，但既没有盗贼闯入，也没有发生地震，同时John和Mary都给你打电话的概率。
- 解：

P(j,m,a,~b,~e) = P(j|a)P(m|a)P(a|~b,~e) P(~b) P(~e)
= 0.9x0.7x0.001x0.999x0.998 = 0.00062
= 0.062%

- 已知，一个事件***e = {JohnCalls = true, and MaryCalls = true}***，试问出现盗贼的概率是多少？
- 解：P(X|e) = αP(X,e) = α$\sum_y$P(X,e,y)

  而P(X,e,y)可写成条件概率乘积的形式。

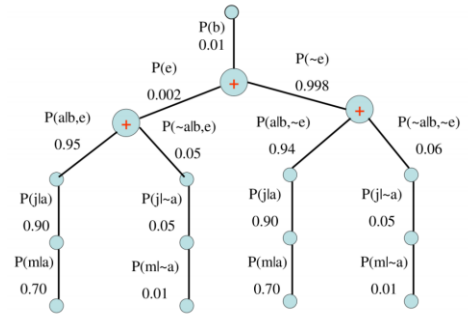
  因此，在贝叶斯网络中可通过计算条件概率的乘积并求和来回答查询。

P(Burgary | JohnCalls = true, MaryCalls = true)简写为：

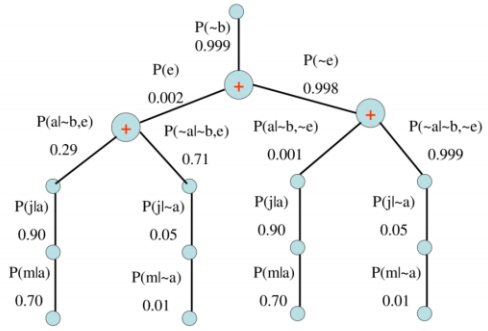
P(B | j, m) = αP(B, j, m) = α $\sum_e\sum_a$P(B, e, a, j, m)

= α $\sum_e\sum_a$ P(b)P(e)P(a|b,e)P(j|a)P(m|a)

= αP(b) $\sum_e$ P(e) $\sum_a$ P(a|b,e)P(j|a)P(m|a)



P(b | j, m)的自顶向下的计算过程



P(~ b | j, m)的自顶向下的计算过程

- P(B | j, m) = αP(B, j, m) = α $\sum_e\sum_a$P(B, e, a, j, m)

= α $\sum_e\sum_a$ P(b)P(e)P(a|b,e)P(j|a)P(m|a)

= αP(b) $\sum_e$ P(e) $\sum_a$ P(a|b,e)P(j|a)P(m|a)

= α×0.001×{[0.002×(0.95×0.9×0.7 + 0.05×0.05 ×0.01)] + [0.998 × (0.94 ×0.9 ×0.7+0.06 ×0.05 ×0.01)]}

= α× 0.00059224

- P(~B | j, m) = αP(~B, j, m) = α $\sum_e\sum_a$P(~B, e, a, j, m)

= α $\sum_e\sum_a$ P(~b)P(e)P(a|~b,e)P(j|a)P(m|a)

= αP(~b) $\sum_e$ P(e) $\sum_a$ P(a|~b,e)P(j|a)P(m|a)

= α×0.999×{[0.002×(0.29×0.9×0.7 + 0.71×0.05 ×0.01)] + [0.998 × (0.001 ×0.9 ×0.7+0.999 ×0.05 ×0.01)]}

= α× 0.0014919

因此，P(B|j, m) = α <0.00059224, 0.0014919>

≈ <0.284, 0.716>

即在John和Mary都打电话的条件下，出现盗贼的概率约为28%。

※类条件概率(条件概率)：假定 x 是一个连续随机变量，其分布取决于类别状态，表示成 p(x|ω)的形式，这就是“类条件概率密度”函数，即类别状态为ω时的 x 的概率密度函数（有时也称为状态条件概率密度）

与**贝叶斯分类器**的联系：**贝叶斯分类器**依据**类条件概率密度 p(x|ω)**和**先验概率 p(ωi)**来判别样本工的类别属性，因此在构建分类器时需要估计出每个类别的先验概率，并且确定类条件概率密度

※连续属性的概率分布计算（此处类比于离散属性）

※示例（连续数据示例）：

根据数据身高 180、体重 120，鞋码 41，请问该人是男是女呢？

编号	身高（CM）	体重（斤）	鞋码（欧码）	性别
1	183	164	45	男
2	182	170	44	男
3	178	160	43	男
4	175	140	40	男
5	160	88	35	女
6	165	100	37	女
7	163	110	38	女
8	168	120	39	女

公式还是上面的公式，这里的困难在于，由于身高、体重、鞋码都是连续变量，不能采用离散变量的方法计算概率。而且由于样本太少，所以也无法分成区间计算。怎么办呢?

这时，可以假设男性和女性的身高、体重、鞋码都是正态分布，通过样本计算出均值和方差，也就是得到正态分布的密度函数。

有了密度函数，就可以把值代入，算出某一点的密度函数的值。

比如，男性的身高是均值 179.5、标准差为 3.697 的正态分布。（我们选择不同条件下的样本，得出的均值，标准差就是条件下的概率分布了。这点稍后计算中体现）

所以男性的身高为 180 的概率为 0.1069。怎么计算得出的呢? –excel

NORMDIST(x,mean,standard_dev,cumulative) 函数，一共有 4 个参数：

. x：正态分布中，需要计算的数值；
. Mean：正态分布的平均值；
. Standard_dev：正态分布的标准差；
. Cumulative：取值为逻辑值，即 False 或 True。它决定了函数的形式。当为 TRUE 时，函数结果为累积分布；为 False 时，函数结果为概率密度。

这里我们使用的是 NORMDIST(180,179.5,3.697,0)=0.1069。

同理我们可以计算得出男性体重为 120 的概率为 0.000382324，

男性鞋码为 41 号的概率为 0.120304111。

P(A1A2A3|C1)=P(A1|C1)P(A2|C1)P(A3|C1)=0.10690.0003823240.120304111=4.9169e-6

同理我们也可以计算出来该人为女的可能性：

条件概率：依据样本点选择

P(A1A2A3|C2)=P(A1|C2)P(A2|C2)P(A3|C2)=0.000001474890.0153541440.120306074=2.7244e-9

2021/6/20 海南大学 计算机科学技术学院

致谢 王沛文 王秀华