31 | 建模连续分布：高斯网络

2018-08-16 王天一

讲述：王天一
时长 14:44 大小 6.76M

无论是贝叶斯网络还是马尔可夫随机场，定义的变量都服从取值有限的离散分布，变量之间的关联则可以用有限维度的矩阵来表示。如果将随机变量的范围从离散型扩展到连续型，变量的可能取值就有无穷多个，这时变量之间的依赖关系就不能再用表格的形式来表示了，需要重新定义概率图模型中的相互作用与条件独立性。

考虑最简单的情形，也就是结点所表示的随机变量都服从高斯分布，**由高斯型连续随机变量构成的概率图模型统称为高斯网络**（Gaussian network）。

如果多个服从一维高斯分布的随机变量构成一个整体，那它们的联合分布就是**多元高斯分布**（multivariate Gaussian distribution），其数学表达式可以写成

$$p(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}}\exp[-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})]$$

其中 $\boldsymbol{\mu}$ 是这组随机变量的均值向量（mean vector），Σ 是这组随机变量的协方差矩阵（covariance matrix），$|\Sigma|$ 是它的行列式值。

协方差矩阵是对称的正定（positive definite）矩阵，表示了不同变量之间的关联：如果两个变量线性无关，那么其协方差矩阵中对应的元素就等于 0，这意味着两个变量满足边际独立性（marginal independency）；如果所有变量都线性无关的话，协方差矩阵就退化为对角矩阵。

协方差矩阵的逆矩阵 $J = \Sigma^{-1}$ 被称为**信息矩阵**（information matrix），信息矩阵和均值向量的乘积则被称为**势向量**（potential vector）。

引入信息矩阵意在定义条件独立性（conditional independency）：和边际独立性不同，条件独立性不能直接在协方差矩阵中体现出来，必须通过信息矩阵加以观察。信息矩阵的元素等于 0 说明对应的两个变量在给定其他变量的前提下条件独立，比如 $J_{1,3} = 0$ 就意味着着在其他变量固定时，x_1 和 x_3 条件独立。

在高斯分布的基础上可以进一步定义高斯线性模型。**高斯线性模型**（linear Gaussian model）指的是一个随机变量可以表示为一组随机变量的线性组合，这个随机变量本身的不确定性则可以用高斯分布来建模，这种关系写成数学表达式就是

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + \epsilon, \epsilon\ \mathcal{N}(0, \sigma^2)$$

这其实和原始线性回归的假设是完全一致的。把这种关系放到概率图模型中，那么所有的 x_i 都可以看成结点 y 的父结点，它们一起构成了汇连结构。从概率角度看，给定这些父结点后，子结点 y 的条件概率就服从高斯分布，其均值是 x_i 的线性组合，方差则是噪声 ϵ 的方差。

上面的表达式中假设所有自变量 x_i 都有固定的取值，如果这些自变量都是随机变量，共同服从均值为 $\boldsymbol{\mu}$，协方差矩阵为 Σ 的多维高斯分布的话，那么可以证明随机变量 y 也是高斯随机变量，它的均值等于 $\beta_0 + \boldsymbol{\beta}^T\boldsymbol{\mu}$，方差等于 $\sigma^2 + \boldsymbol{\beta}^T\Sigma\boldsymbol{\beta}$，和变量 X_i 的协方差则等于 $\sum_{j=1}^{k}\beta_j\Sigma_{i,j}$。

这样的结论告诉我们，**高斯线性模型实际上定义了一个高斯贝叶斯网络**（Gaussian Bayesian network），**整个概率图所表示的联合分布就是一个大的多维高斯分布。**

高斯贝叶斯网络的表示可以用下面这个例子来直观地解释，这个例子来自《概率图模型》（Probabilistic Graphical Models）的例 7.3。

“如果一个线性高斯网络具有顺连结构 $X_1 \to X_2 \to X_3$，其中 X_1 的概率密度 $\mathcal{N}(1,4)$，已知 X_1 时 X_2 的条件概率密度为 $\mathcal{N}(0.5X_1 - 3.5, 4)$，已知 X_2 时 X_3 的条件概率密度为 $\mathcal{N}(-X_2 + 1, 3)$，试求解整个网络所表示的联合分布。 “

在高斯形式已经确定的前提下，求解联合分布实际上就是求解所有变量的均值向量和协方差矩阵。由于 X_2 等于 $0.5X_1 - 3.5$，将 X_1 的均值为 1 代入这个线性关系，就可以求出 X_2 的均值等于 $0.5 \times 1 - 3.5 = -3$，同理可以求出 X_2 的均值等于 $-(-3) + 1 = 4$。

求完了均值再来看协方差，协方差矩阵的对称性决定了对于 3 维变量来说，计算协方差矩阵需要确定 6 个元素。X_1 的方差 $\Sigma_{11} = 4$ 是已知的，这部分方差将会以线性系数为比例体现在 X_2 中，和 X_2 自身的不确定度共同构成随机变量完整的方差，也就是 $\Sigma_{22} = 0.5^2 \times 4 + 4 = 5$。将 X_2 的方差代入 X_3 的线性关系，又可以计算出 $\Sigma_{33} = (-1)^2 \times 5 + 3 = 8$。这三个方差定义了变量自身的不确定性，是协方差矩阵中的对角线元素。

确定了对角线元素后，下一步就是确定非对角线上的元素，也就是不同变量之间的相关性。由于 X_2 这个变量只取决于 X_1，其关联的强度由线性系数确定，因而两者之间的协方差就等于线性系数和 X_1 方差的乘积 $\Sigma_{12} = 0.5 \times \Sigma_{11} = 2$。这个数字的含义在于用 X_1 的变化对 X_2 的变化的影响。同理可以求出，X_2 和 X_3 之间的协方差为 $\Sigma_{23} = -1 \times \Sigma_{22} = -5$。

在这个顺连结构中，X_1 和 X_3 之间并不存在直接的作用，而是以 X_2 作为媒介和中转。X_1 对 X_3 的作用实际上可以分成两个阶段，第一个阶段是 X_1 的变化首先影响 X_2，第二个阶段是 X_2 的变化继续影响 X_3。在协方差的计算中，第一个阶段体现为 X_1 和 X_2 之间的协方差，第二个阶段则体现为 X_2 和 X_3 之间的线性系数的加权作用。两者相乘形成了一个整体，也就是 $\Sigma_{13} = \Sigma_{12} \cdot (-1) = -2$。由此，就可以写出联合分布的均值向量和协方差矩阵

$$\boldsymbol{\mu} = (1, -3, 4)^T, \Sigma = \begin{pmatrix} 4 & 2 & -2 \\ 2 & 5 & -5 \\ -2 & -5 & 8 \end{pmatrix}$$

关于这个例子需要说明的一点是，由于协方差矩阵中所有的元素都不为 0，说明这些变量两两之间都不是边际独立的。但顺连结构告诉我们，当 X_2 确定时，X_1 和 X_3 条件独立，所以它的信息矩阵中会有两个零元素。这说明在图结构中，表示同一个联合分布只需要更少的参数。

但淮南为橘淮北为枳，图结构的优势也可能变成劣势。想象一下汇连结构 $X_1 \rightarrow X_2 \leftarrow X_3$，由于汇连结构中不存在条件独立的结点，因此联合分布的信息矩阵中所有元素都是非零的。但由于 X_1 和 X_3 互不影响，因此协方差矩阵中反倒存在着零元素。

在此基础上，如果再给结点 X_1 和 X_3 赋予一个共同的父结点 X_4，让这三者形成分连结构的话，那整个网络中就既没有条件独立性，也没有边际独立性，无论是协方差矩阵还是信息矩阵中就都不会出现非零元素了。

将多元高斯分布嵌入到无向的马尔可夫随机场中，得到的就是高斯马尔可夫随机场（Gaussian Markov random field）。在处理高斯随机场时，先要对多元高斯分布的概率密度做些处理，将指数项中的协方差逆矩阵 Σ^{-1} 替换为信息矩阵 J 并展开。由于均值向量和信息矩阵都是常量，将它们去掉就可以得到概率密度的正比关系

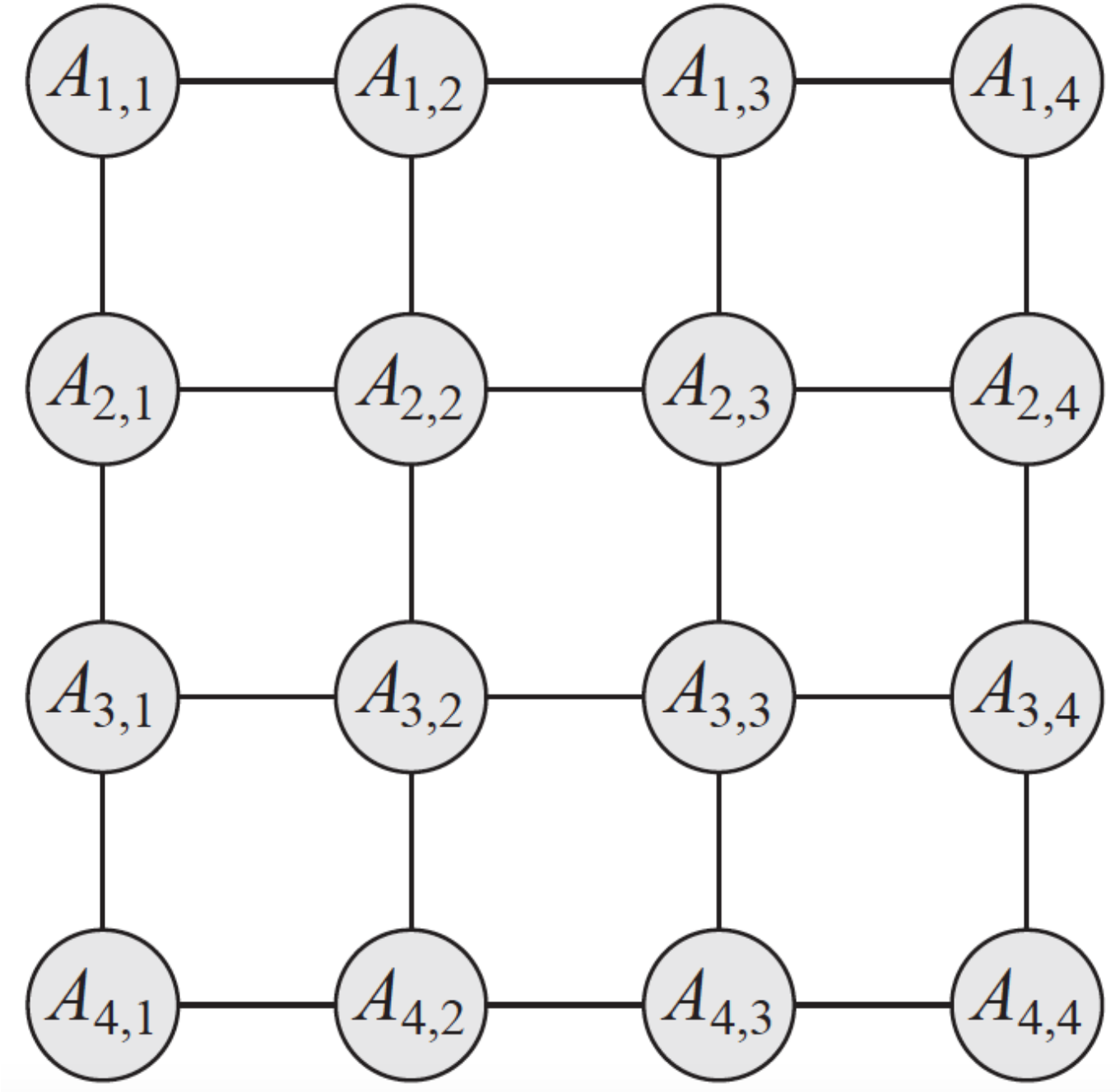
$$p(\mathbf{s}) \propto \exp[-\frac{1}{2}\mathbf{x}^T J\mathbf{x} + (J\boldsymbol{\mu})^T\mathbf{x}]$$

这个式子被称为**高斯分布的信息形式**（information form）。由于式子中的 $\mathbf{x}$ 是向量，因此展开后的结果中会包含两种多项式成分：一种成分是单个变量 X_i 的函数，其表达式可以写成 $-J_{i,i}x_i^2/2 + h_i x_i$，其中 h_i 是势向量的第 i 个分量；另一种成分是两个变量 X_i 和 X_j 乘积的函数，其表达式可以写成 $-J_{i,j}x_i x_j$。

在高斯随机场中，这两个不同的成分具有不同的意义。只与单个变量相关的成分可以看成每个结点的势函数（node potential），同时涉及两个变量的成分则可以看成连接这两个结

点的边的势函数（edge potential）。如果信息矩阵的元素 $J_{i,j} = 0$，其对应的边势也等于 0，就说明这两个结点之间并没有连接的边。

需要说明的是，在前一篇对马尔可夫随机场的介绍中我提到了边势，但并没有涉及结点势的概念。其原因在于结点势并不是通用的概念，它只存在于具有成对马尔可夫性的网络之中。下图是一个典型的成对马尔可夫随机场，每个结点都和它所有的非邻接结点条件独立，在信息矩阵 J 中，这些条件独立的结点组合所对应的元素就等于 0。



成对马尔可夫随机场（图片来自 Probabilistic Graphical Models，图 4.A.1）

多元高斯分布定义的是成对的马尔可夫随机场，其中的每个势函数都具有二次型的形式。反过来，由于任何合法的高斯分布都具有正定的信息矩阵，所以如果一个成对随机场能够改写成多元高斯分布，那它的势函数的系数所形成的矩阵也必须得满足正定的条件。

对连续分布的建模能够大大拓展概率图模型的应用范围，毕竟现实中大量的观测结果都是连续变化的。**虽然高斯分布并不适用于所有的连续变量，但良好的数学性质和便于计算的特点让它成为了理想条件下近似建模的首选。**

如果一个概率图模型中的随机变量既有离散型也有连续型，这样的网络就是**混合网络**（hybrid network）。混合网络让人头疼的一个问题是同一个结点的父结点可能存在不同的类型，其中既有连续分布的结点，也有离散分布的结点。而在处理这些父结点不同的子结点时，需要根据情况分类讨论。

如果子结点是连续分布的结点，那么问题就简单了。由于离散分布的父结点取值的组合是有限的，就可以对每一种可能的取值都为子结点定义一组线性系数，将离散结点的信息编码到这组线性系数当中。

这样一来，子结点就可以表示成连续父结点的线性组合，其中的线性系数则由离散父结点来决定。这种模型被称为**条件线性模型**（conditional linear model），它本质上是一组不同参数的高斯分布所形成的混合模型（mixture model），每个分布的权重取决于这一组参数出现的概率。

当一个离散的子结点具有连续的父结点时，解决的方法也不复杂。最简单的办法是采用**阈值模型**（threshold model），当连续变量的取值大于阈值时输出 1，小于阈值时输出 0。更加精细的一种方式是用借鉴逻辑回归或者 softmax 回归的思想，计算离散的子结点关于连续的父结点的条件概率，并输出条件概率最大的结果。

今天我和你分享了概率图模型中对连续型随机变量的建模与表示，其要点如下：

高斯网络采用高斯线性模型建模连续变量，其数字特征为均值向量和协方差矩阵；

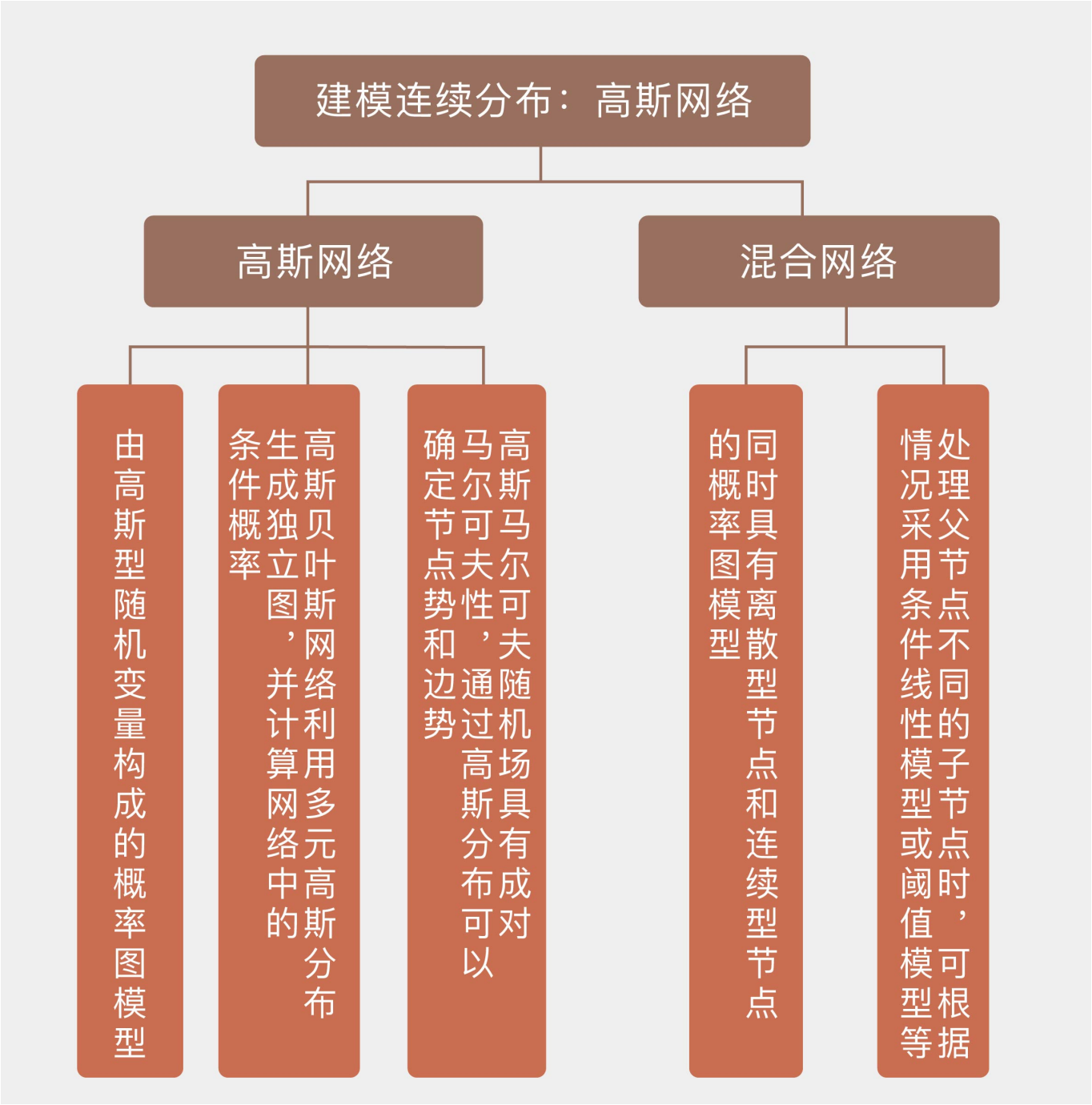
高斯贝叶斯网络利用多元高斯分布生成独立图，利用信息矩阵计算网络中的条件概率；

高斯马尔可夫随机场具有成对马尔可夫性，通过高斯分布可以确定结点势和边势；

混合网络是同时具有离散型结点和连续型结点的概率图模型。

在现实生活中，自然界客观存在的属性通常是连续分布的，而人为定义出来的属性则通常是离散的，那么你能想象出有哪些离散变量和连续变量共存的应用场景呢?

欢迎分享你的观点。



上一篇　30 | 无向图模型：马尔可夫随机场

下一篇　32 | 从有限到无限：高斯过程

精选留言 (1)

写留言

林彦

2018-09-09

比如在一个网上商城的商品，它既会有类别这些离散变量，也会有价格等连续变量，然后要组合在一起放入模型中预测。