斯坦福 ML 公开课 13B

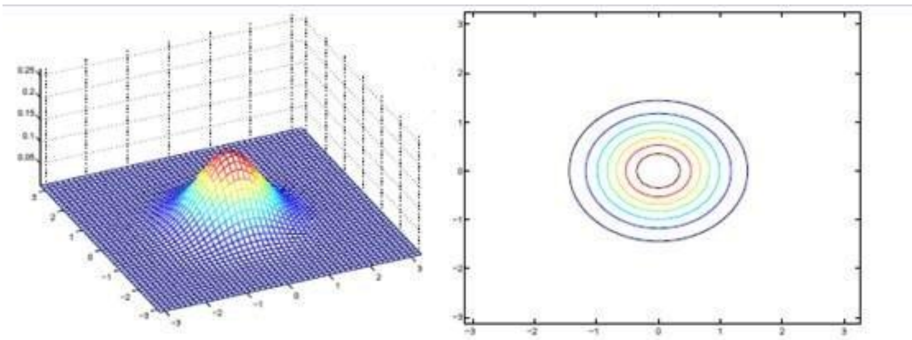
本文是《斯坦福 ML 公开课笔记 13A》的续篇。主要讲述针对混合高斯模型的问题所采取的简单解决方法，即对假设进行限制的简单方法，最后引出因子分析模型（Factor Analysis Model），包括因子分析模型的介绍、EM 求解等。

混合高斯模型的问题

在上一篇笔记中，谈到对于混合高斯模型来说，当训练数据样本数目小于样本的维度时，因为协方差矩阵的非奇异性，导致不能得到概率密度函数的问题。而对于其他模型来说，样本数小于样本维度，也容易引起过拟合的问题。

追本溯源，这个问题可以认为是数据信息缺乏的问题，即从训练数据中得不到模型所需的全部信息。解决办法就是减少模型所需要的信息，本文提到的手段有两个，第一个就是不改变现有模型，但是加强模型的假设，下面提到的对协方差矩阵的限制即是此类。第二个手段则是降低模型的复杂度，提出一个需要更少参数（更少参数即是需要更少信息）的模型，因子分析模型即是此类。

限制协方差矩阵的方法，其实在 13A 文章中已提到。一个稍弱的假设是假设协方差矩阵为对角矩阵，更强的假设是假设协方差矩阵为对角矩阵且对角线上的值都相等。怎样直观的理解这两个假设呢？请看下图。



对于二维多元高斯分布来说，它有一个几何特性。即在平面上的投影是一个椭圆，当假设该分布的协方差矩阵为对角矩阵时，那么这个椭圆的轴就与坐标轴平行。当限制对角线上的值都相等时，那么投影就变成了圆。

当需要估计出完整的协方差矩阵时，需要的样本数目 m 必须大于样本维度 n。但是当有上述对角线假设时，只要样本数目大于 1 就可以估计出限定的协方差矩阵。

接下来讨论因子分析模型，在介绍因子分析模型之前，先看高斯分布的另一种写法，该写法是推导因子分析模型的基础。

高斯分布的矩阵写法

下面我们先看高斯分布的另一种写法。假设我们拥有三个随机向量$x_1 \in \mathbb{R}^r, x_2 \in \mathbb{R}^s$。

$$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \tag{1}$$

那么$x \in \mathbb{R}^{r+s}$，假设$x \sim \mathcal{N}(\mu, \Sigma)$，且

$$\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \tag{2}$$

这里，$\mu_1 \in \mathbb{R}^r, \mu_2 \in \mathbb{R}^s, \Sigma_{11} \in \mathbb{R}^{r*r}, \Sigma_{12} \in \mathbb{R}^{r*s}, \Sigma_{21} \in \mathbb{R}^{s*r}, \Sigma_{22} \in \mathbb{R}^{s*s}$。因为协方差矩阵

是对称的，因而$\Sigma_{12} = (\Sigma_{12})^T$。

在这些前提下，考虑如何求得x_1的边际分布？由公式 1 和公式 2 不难看到：

$$\mathrm{E}[x_1] = \mu_1 \tag{3}$$

$$\mathrm{Cov}(x_1) = \mathrm{E}[(x_1 - \mu_1)(x_1 - \mu_1)^T] = \Sigma_{11} \tag{4}$$

当然，可以简单的对公式 2 推导一下：

$$\mathrm{Cov}(\mathrm{x}) = \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} = \mathrm{E}[(x-\mu)(x-\mu)^T]$$

$$= \mathrm{E}\left[\begin{pmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \end{pmatrix}\begin{pmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \end{pmatrix}^T\right] = \mathrm{E}\begin{bmatrix} (x_1-\mu_1)(x_1-\mu_1)^T & (x_1-\mu_1)(x_2-\mu_2)^T \\ (x_2-\mu_2)(x_1-\mu_1)^T & (x_2-\mu_2)(x_2-\mu_2)^T \end{bmatrix} \tag{5}$$

从而，我们知道，$x_1 \sim \mathcal{N}(\mu_1, \Sigma_{11})$。

那么另一个问题，在给定x_2时的x_1的条件概率是什么？

$$\mathrm{p}(x_1|x_2) = \frac{p(x_1, x_2)}{p(x_2)} = \frac{p(x)}{p(x_2)} \tag{6}$$

因为，$x \sim \mathcal{N}(\mu, \Sigma)$，$x_2 \sim \mathcal{N}(\mu_2, \Sigma_{22})$，两个正态分布相除得到一个新的正态分布。推导过程我也没推出来，不过据说 Chuong B. Do 写的《Gaussian processes》中有此项推导，容后再写吧。这里直接写结果了。

$$x_1|x_2 \sim \mathrm{N}(\mu_{1|2}, \Sigma_{1|2}) \tag{7}$$

其中，

$$\mu_{1|2} = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2) \tag{8}$$

$$\Sigma_{1|2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} \tag{9}$$

因子分析模型

因子分析模型的定义如下：

假设有隐含变量$z \sim \mathcal{N}(0, \mathrm{I})$，$\mathrm{z} \in \mathbb{R}^d, (\mathrm{d} < \mathrm{n})$。

再假设训练样本 x 由隐含变量 z 生成，即

$$\mathrm{x} = \mu + \lambda \mathrm{z} + \varepsilon \tag{10}$$

其中，$\varepsilon \sim \mathrm{N}(0, \psi)$.

公式 10 等价于当 z 已知的时候，x 的概率分布，如公式 11 所示:

$$\mathrm{x}|\mathrm{z} \sim \mathrm{N}(\mu + \lambda \mathrm{z}, \psi) \tag{11}$$

这即是因子分析模型的定义，该模型有三个参数，$\mu \in \mathbb{R}^n$，$\lambda \in \mathbb{R}^{n \times d}$, $\psi \in \mathbb{R}^{n \times n}$，$\psi$是对角矩阵。

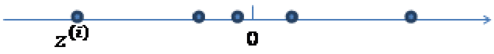
因子分析模型可以从训练数据生成过程上进行理解：

1) 首先，在一个低维空间内用均值为 0，协方差为单位矩阵的多元高斯分布生成 m 个隐含变量$\mathrm{z}^{(\mathrm{i})}$，$\mathrm{z}^{(\mathrm{i})}$是 d 维向量，m 也是样本数目。
2) 然后使用变换矩阵λ将 z 映射到 n 维空间λz。此时因为 z 的均值为 0，映射后的均值仍然为 0。
3) 再然后将 n 维向量λz再加上一个均值μ,对应的意义是将变换后的 z 的均值在 n 维空间上平移。
4) 由于真实样例 x 会有误差，在上述变换的基础上再加上误差$\varepsilon \sim \mathrm{N}(0, \psi)$
5) 最后的结果是认为训练样例的生成公式为

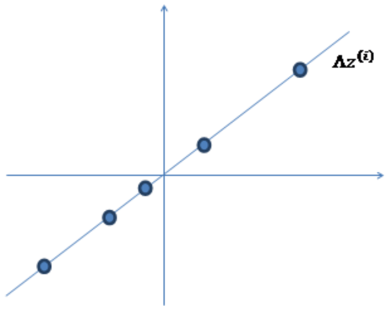
$$\mathrm{x} = \mu + \lambda \mathrm{z} + \varepsilon$$

在视频中，Ng 举出了一个生成样本的例子来方便大家理解因子分析模型，假设$\mathrm{z} \in \mathbb{R}^1, \mathrm{x} \in \mathbb{R}^2$。z 是一维向量，x 是二维向量。再假设$\lambda = [1\ 2]^T, \psi = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}, \mu = [3\ 1]^T$。

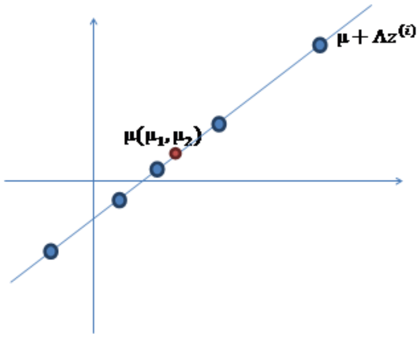
按照生成过程的 5 步，第一步，生成 m 个隐含变量。



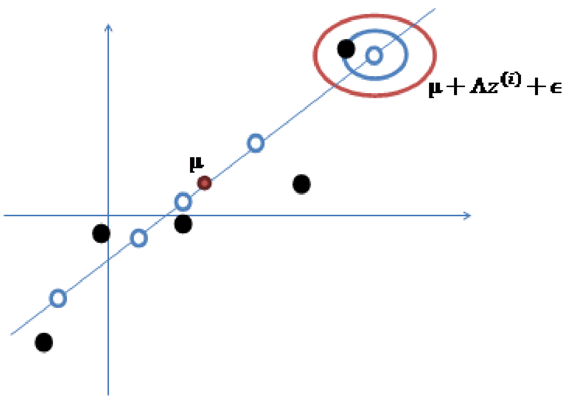
第二步，使用λ转换维度。



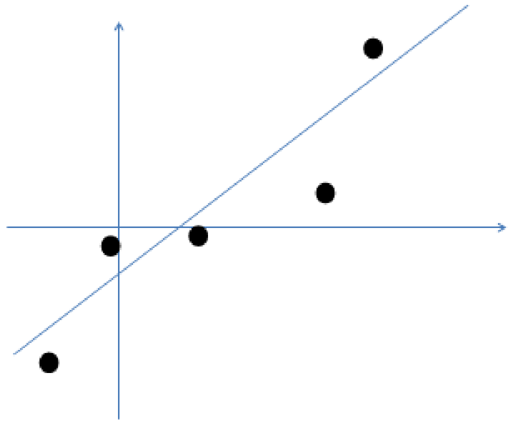
第三步，使用μ进行平移。



第四步，加入随机扰动。



第五步，得到最终的训练数据。



为了更方便大家的理解，在此举一个实际中使用的因子分析模型的例子。

在企业形象或品牌形象的研究中，消费者可以通过一个有 24 个指标构成的评价体系，评价百货商场的 24 个方面的优劣。但消费者主要关心的是三个方面，即商店的环境、商店的服务和商品的价格。因子分析方法可以通过 24 个变量，找出反映商店环境、商店服务水平和商品价格的三个潜在的因子，对商店进行综合评价。

因子分析模型的推导

上一节对因子分析模型进行了定义，以及从数据生成的角度对它进行了进一步阐述。本节则介绍上一节中定义的参数在模型中是如何被使用的。具体来讲，就是该模型针对训练数据的似然函数是什么。也就是说，上一节讲述了因子是什么，本节讲述如何分析。

首先，重新列出模型的定义公式。

$$z\sim\mathcal{N}(0,\mathrm{I})$$
$$\varepsilon\sim\mathrm{N}(0,\psi)$$
$$\mathrm{x}=\mu+\lambda\mathrm{z}+\varepsilon$$

其中，误差ε和隐含因子 z 是相互独立的。

使用高斯分布的矩阵表示法对模型进行分析。该方法认为 z 和 x 符合多元高斯分布，即：

$$\begin{bmatrix} z \\ x \end{bmatrix}\sim N(\mu_{zx},\Sigma) \tag{12}$$

接下来就是求解μ_{zx},Σ。

已知 E[z]=0，E[ε]=0 则

$$\mathrm{E[x]}=\mathrm{E}[\mu+\lambda\mathrm{z}+\varepsilon]=\mu \tag{13}$$

所以

$$\mu_{zx}=\begin{bmatrix} \vec{0} \\ \mu \end{bmatrix} \tag{14}$$

为了求解Σ，需要计算$\Sigma_{zz}=E[(z-E[z])(z-E[z])^T]$, $\Sigma_{zx}={\Sigma_{xz}}^T=E[(z-E[z])(x-E[x])^T]$和$\Sigma_{\mathrm{xx}}=E[(x-E[x])(x-E[x])^T]$。

由定义，可知$\Sigma_{zz}=\mathrm{Cov(z)}=\mathrm{I}$,另外，还有

$$E[(z-E[z])(x-E[x])^T]=\mathrm{E}[\mathrm{z}(\mu+\lambda\mathrm{z}+\varepsilon-\mu)^T]=\mathrm{E}[\mathrm{zz}^T]\lambda^T+E[z\varepsilon^T]=\lambda^T$$

上述公式的最后一步，$\mathrm{E}[\mathrm{zz}^T]=\mathrm{Cov(z)}=\mathrm{I}$, z 和$\varepsilon$相互独立，也有$E[z\varepsilon^T]=E[z]E[\varepsilon^T]=0$。

$$\begin{aligned}E[(x-E[x])(x-E[x])^T]&=\mathrm{E}[(\lambda\mathrm{z}+\varepsilon)(\lambda\mathrm{z}+\varepsilon)^T]\\&=\mathrm{E}[\lambda\mathrm{zz}^T\lambda^T+\varepsilon\mathrm{z}^T\lambda^T+\lambda\mathrm{z}\varepsilon^T+\varepsilon\varepsilon^T]=\lambda\mathrm{E}[\mathrm{zz}^T]\lambda^T+E[\varepsilon\varepsilon^T]=\lambda\lambda^T+\psi\end{aligned} \tag{15}$$

将上述求解结果放到一起，得到

$$\begin{bmatrix} z \\ x \end{bmatrix} \sim N(\begin{bmatrix} \vec{0} \\ \mu \end{bmatrix}, \begin{bmatrix} I & \lambda^T \\ \lambda & \lambda\lambda^T + \psi \end{bmatrix}) \tag{16}$$

所以，我们得到 x 的边际分布为

$$x \sim \mathcal{N}(\mu, \lambda\lambda^T + \psi) \tag{17}$$

因而，对于一个训练集$\{x^{(i)}; \mathrm{i} = 1,2, \dots, \mathrm{m}\}$，我们可以写出参数的似然函数。

$$\ell(\mu, \lambda, \psi) = \log \prod_{i=1}^{m} \frac{1}{(2\pi)^{n/2} |\lambda\lambda^T + \psi|^{\frac{1}{2}}} \times$$

$$exp\left(-\frac{1}{2}(x^{(i)} - \mu)(\lambda\lambda^T + \psi)^{-1}(x^{(i)} - \mu)^T\right) \tag{18}$$

由上式，若是直接最大化似然函数的方法求解参数的话，你会发现很难，因而下一节会介绍使用 EM 算法求解因子分析的参数。

EM 求解参数

同 13A 中 MoG 和 MoNB 模型一样，因子分析模型的 EM 求解也是直接套 EM 一般化算法中的 E-step 和 M-step 中的公式。对于 E-step 来说，

$$Q_i(z^{(i)}) = p(z^{(i)} | x^{(i)}; \mu, \lambda, \psi) \tag{19}$$

在高斯分布的矩阵写法一节中，我们已经算出了条件概率的期望和方差分别是什么了，如下所示

$$\mu_{z^{(i)}|x^{(i)}} = \lambda^T(\lambda\lambda^T + \psi)^{-1}(x^{(i)} - \mu) \tag{20}$$

$$\Sigma_{z^{(i)}|x^{(i)}} = I - \lambda^T(\lambda\lambda^T + \psi)^{-1}\lambda \tag{21}$$

代入上面两个公式，就可以得到$Q_i(z^{(i)})$的概率密度函数了，即：

$$Q_i(z^{(i)}) = \frac{1}{(2\pi)^{n/2} |\Sigma_{z^{(i)}|x^{(i)}}|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(x^{(i)} - \mu_{z^{(i)}|x^{(i)}}) {\Sigma_{z^{(i)}|x^{(i)}}}^{-1} (x^{(i)} - \mu_{z^{(i)}|x^{(i)}})^T\right) \tag{22}$$

在 M-step 中，需要最大化如下公式来求取参数μ, λ, ψ。

$$\sum_{i=1}^{m} \int Q_i(z^{(i)}) log \frac{p(z^{(i)}, x^{(i)}; \mu, \lambda, \psi)}{Q_i(z^{(i)})} dz^{(i)}$$

$$= \sum_{i=1}^{m} \int Q_i(z^{(i)}) [logp(x^{(i)} | z^{(i)}; \mu, \lambda, \psi) + logp(z^{(i)}) - logQ_i(z^{(i)})] dz^{(i)}$$

$$= \sum_{i=1}^{m} E_{z^{(i)} \sim Q_i} [logp(x^{(i)} | z^{(i)}; \mu, \lambda, \psi) + logp(z^{(i)}) - logQ_i(z^{(i)})] \tag{23}$$

上面公式中，第一步转化是先将 p(z,x)=p(x|z)*p(z)，然后将 log 打开。第二部则是将积分转化为求 z 服从 Q 分布的时候，函数$logp(x^{(i)} | z^{(i)}; \mu, \lambda, \psi) + logp(z^{(i)}) - logQ_i(z^{(i)})$的期望。本文下面会省略 E 的下标。

下文以对λ求解为例，对公式进行求解。首先，对目标函数进行简化。

$$\nabla_\lambda \sum_{i=1}^{m} E[logp(x^{(i)} | z^{(i)}; \mu, \lambda, \psi) + logp(z^{(i)}) - logQ_i(z^{(i)})]$$

$$=\nabla_{\lambda}\sum_{i=1}^{m}E[logp(x^{(i)}|z^{(i)};\mu,\lambda,\psi)]$$

$$=\nabla_{\lambda}\sum_{i=1}^{m}\mathrm{E}[\frac{1}{(2\pi)^{n/2}|\psi|^{\frac{1}{2}}}\exp\left(-\frac{1}{2}(x^{(i)}-\mu-\lambda z^{(i)})\psi^{-1}(x^{(i)}-\mu-\lambda z^{(i)})^{T}\right)]$$

$$=\nabla_{\lambda}\sum_{i=1}^{m}E[-\frac{1}{2}\log|\psi|-\frac{n}{2}\log(2\pi)-\frac{1}{2}(x^{(i)}-\mu-\lambda z^{(i)})\psi^{-1}(x^{(i)}-\mu-\lambda z^{(i)})^{T}]$$

$$=\nabla_{\lambda}\sum_{i=1}^{m}-\mathrm{E}[\frac{1}{2}(x^{(i)}-\mu-\lambda z^{(i)})\psi^{-1}(x^{(i)}-\mu-\lambda z^{(i)})^{T}]$$

$$=\sum_{i=1}^{m}\nabla_{\lambda}\mathrm{E}[-\mathrm{tr}\frac{1}{2}z^{(i)^{T}}\lambda^{T}\psi^{-1}\lambda z^{(i)}+\mathrm{tr}z^{(i)^{T}}\lambda^{T}\psi^{-1}(x^{(i)}-\mu)]$$

$$=\sum_{i=1}^{m}\nabla_{\lambda}\mathrm{E}[-\mathrm{tr}\frac{1}{2}\lambda^{T}\psi^{-1}\lambda z^{(i)}z^{(i)^{T}}+\mathrm{tr}\lambda^{T}\psi^{-1}(x^{(i)}-\mu)z^{(i)^{T}}]$$

$$=\sum_{i=1}^{m}\left(\nabla_{\lambda}\mathrm{E}\left[-\mathrm{tr}\frac{1}{2}\lambda^{T}\psi^{-1}\lambda z^{(i)}z^{(i)^{T}}\right]+\nabla_{\lambda}\mathrm{E}\left[\mathrm{tr}\lambda^{T}\psi^{-1}\left(x^{(i)}-\mu\right)z^{(i)^{T}}\right]\right)$$

$$=\sum_{i=1}^{m}\left(\mathrm{E}\left[-\nabla_{\lambda}\mathrm{tr}\frac{1}{2}\lambda^{T}\psi^{-1}\lambda z^{(i)}z^{(i)^{T}}\right]+\mathrm{E}\left[\nabla_{\lambda}\mathrm{tr}\lambda^{T}\psi^{-1}\left(x^{(i)}-\mu\right)z^{(i)^{T}}\right]\right)$$

$$=\sum_{i=1}^{m}\left(\mathrm{E}\left[-(\nabla_{\lambda^{T}}\mathrm{tr}\frac{1}{2}\lambda^{T}\psi^{-1}\lambda z^{(i)}z^{(i)^{T}})^{T}\right]+\mathrm{E}\left[(\nabla_{\lambda^{T}}\mathrm{tr}\lambda^{T}\psi^{-1}\left(x^{(i)}-\mu\right)z^{(i)^{T}})^{T}\right]\right)$$

$$=\sum_{i=1}^{m}\left(\mathrm{E}\left[-\frac{1}{2}(2z^{(i)}z^{(i)^{T}}\lambda^{T}\psi^{-1})^{T}\right]+\mathrm{E}\left[((\psi^{-1}\left(x^{(i)}-\mu\right)z^{(i)^{T}})^{T})^{T}\right]\right)$$

$$=\sum_{i=1}^{m}\left(\mathrm{E}\left[-\psi^{-1}\lambda z^{(i)}z^{(i)^{T}}\right]+\mathrm{E}\left[\psi^{-1}\left(x^{(i)}-\mu\right)z^{(i)^{T}}\right]\right)$$

$$=\sum_{i=1}^{m}\mathrm{E}\left[-\psi^{-1}\lambda z^{(i)}z^{(i)^{T}}+\psi^{-1}\left(x^{(i)}-\mu\right)z^{(i)^{T}}\right] \tag{24}$$

该简化过程分为 12 步，较长，下面一步一步进行解析。第一步，去除与参数λ无关的项。第二步，将$p(x^{(i)}|z^{(i)};\mu,\lambda,\psi)$的概率密度函数代入，$p(x^{(i)}|z^{(i)};\mu,\lambda,\psi)$的期望为$\mu+\lambda z^{(i)}$，方差为$\psi$。第三步，将概率密度函数用 log 打开。第四步，去除与参数λ无关的项。第五步，将三项连乘打开，并提取与参数λ相关的项，然后由于最后提取得到的两项的乘积结果都是实数，利用矩阵的迹的性质 tr(a)=a，使用迹替换。第六步，利用矩阵的迹的性质 tr(AB)=tr(BA)，将$z^{(i)^{T}}$项放到最后。第七步，将期望打开。第八步，将求导换到期望里面去，因为期望是针对 z，求导是针对λ，所以切换是可以的。第九步，利用矩阵的迹的性质$\nabla_{A^{T}}\mathrm{f}(A)=(\nabla_{A}f(A))^{T}$，

将求导符下标λ代换为λ^{T}，两项都是这样。第十步，对于第一项，利用矩阵$\nabla_A trABA^TC = CAB + C^TAB^T$的性质代入，对于第二项，将矩阵$\nabla_A \mathrm{tr}AB = B^T$的性质代入。第十一步，将转置符号打开代入，需要说明的是，对于ψ^{-1}来说，由于其是对角矩阵，所以转置与自身相等。第十二步，期望归并到一起。矩阵的迹的性质，可以参考本系列笔记 1-2。

将最后的结果设为 0 并简化，可得

$$\sum_{i=1}^{m} \lambda E_{z^{(i)}\sim Q_i}\left[\mathrm{z}^{(\mathrm{i})}\mathrm{z}^{(\mathrm{i})^{\mathrm{T}}}\right] = \sum_{i=1}^{m} (\mathrm{x}^{(\mathrm{i})} - \mu) E_{z^{(i)}\sim Q_i}[\mathrm{z}^{(\mathrm{i})^{\mathrm{T}}}] \tag{25}$$

所以，可以求解λ

$$\lambda = (\sum_{i=1}^{m} (\mathrm{x}^{(\mathrm{i})} - \mu) E_{z^{(i)}\sim Q_i}[\mathrm{z}^{(\mathrm{i})^{\mathrm{T}}}])(\sum_{i=1}^{m} E_{z^{(i)}\sim Q_i}\left[\mathrm{z}^{(\mathrm{i})}\mathrm{z}^{(\mathrm{i})^{\mathrm{T}}}\right])^{-1} \tag{26}$$

到这里我们发现，公式 26 与回归中的最小二乘法的矩阵形式类似：

$$\theta^T = (y^TX)(X^TX)^{-1}$$

在因子分析模型中，x 是 z 的线性函数，在 E-step 中给出 z 的猜测值后，在 M-step 中寻找 x 和 z 之间的映射关系λ。而最小二乘法也是寻找 x 与 y 之间的线性关系，所以它们之间才会相似。它们之间的区别在于，最小二乘法只用到了 z 的最优估计，而因子分析还用到了$\mathrm{z}^{(\mathrm{i})}\mathrm{z}^{(\mathrm{i})^{\mathrm{T}}}$的估计。

到这λ的求解公式 26 中还有未知量。下面一一求解。

因为Q_i的定义如公式 19 所示。所以可以很容易发现

$$E_{z^{(i)}\sim Q_i}\left[\mathrm{z}^{(\mathrm{i})^{\mathrm{T}}}\right] = \mu_{z^{(i)}|x^{(i)}}^T \tag{27}$$

对于一个随机变量 Y 来说，有一个性质$\mathrm{Cov(Y) = E[YY}^T] - \mathrm{E[Y]E[Y}^T]$,所以有

$$E_{z^{(i)}\sim Q_i}\left[\mathrm{z}^{(\mathrm{i})}\mathrm{z}^{(\mathrm{i})^{\mathrm{T}}}\right] = \mu_{z^{(i)}|x^{(i)}}\mu_{z^{(i)}|x^{(i)}}^T + \Sigma_{z^{(i)}|x^{(i)}} \tag{28}$$

将公式 27、28 代入进公式 26，即可得到最终的求解公式。

$$\lambda = (\sum_{i=1}^{m} (\mathrm{x}^{(\mathrm{i})} - \mu)\mu_{z^{(i)}|x^{(i)}}^T)(\sum_{i=1}^{m} \mu_{z^{(i)}|x^{(i)}}\mu_{z^{(i)}|x^{(i)}}^T + \Sigma_{z^{(i)}|x^{(i)}})^{-1} \tag{29}$$

注意到公式中使用到了$\Sigma_{z^{(i)}|x^{(i)}}$，它是后验概率 p(z|x)的方差，在 M-step 中必须得考虑到 z 的后验概率。EM 中一个常见的错误是在 E-step 中只需计算隐含变量 z 的期望 E[z]，而后在 m-step 中的每个 z 出现的地方代入。这样做在 MoG 和 MoNB 中可用，但在因子分析中还需要计算$E[\mathrm{zz^T}]$，因为必须得将 z 的后验分布 p(z|x)考虑进来。

最后，使用同样的方法（求导取 0），在 M-step 中还可以求得另外两个参数μ, ψ，篇幅有限，本文只给出结果。

$$\mu = \frac{1}{m}\sum_{i=1}^{m} x^{(i)} \tag{30}$$

$$\phi = \frac{1}{m}\sum_{i=1}^{m} x^{(i)}x^{(i)^T} - x^{(i)}\mu_{z^{(i)}|x^{(i)}}^T\lambda^T - \lambda\mu_{z^{(i)}|x^{(i)}}x^{(i)^T} + \lambda(\mu_{z^{(i)}|x^{(i)}}\mu_{z^{(i)}|x^{(i)}}^T + \Sigma_{z^{(i)}|x^{(i)}})\lambda^T \tag{31}$$

注意到，上式中求解的不是ψ而是ϕ，在计算出ϕ之后还需再来一步，将$\psi_{\mathrm{ii}} = \phi_{ii}$，因为求解出的ϕ不是对角矩阵，所以只需取ϕ的对角线上的值即可。