

因子分析 (Factor Analysis)

JerryLead@ISCAS

csxulijie@gmail.com

2011年5月11日

1 问题

之前我们考虑的训练数据中样例 $x^{(i)}$ 的个数 m 都远远大于其特征个数 n , 这样不管是进行回归、聚类等都没有太大的问题。然而当训练样例个数 m 太小, 甚至 $m \ll n$ 的时候, 使用梯度下降法进行回归时, 如果初值不同, 得到的参数结果会有很大偏差 (因为方程数小于参数个数)。另外, 如果使用多元高斯分布(Multivariate Gaussian distribution)对数据进行拟合时, 也会有问题。让我们来演算一下, 看看会有什么问题:

多元高斯分布的参数估计公式如下:

$$\mu = \frac{1}{m} \sum_{i=1}^m x^{(i)}$$
$$\Sigma = \frac{1}{m} \sum_{i=1}^m (x^{(i)} - \mu)(x^{(i)} - \mu)^T$$

分别是求 mean 和协方差的公式, $x^{(i)}$ 表示样例, 共有 m 个, 每个样例 n 个特征, 因此 μ 是 n 维向量, Σ 是 $n \times n$ 协方差矩阵。

当 $m \ll n$ 时, 我们会发现 Σ 是奇异阵 ($|\Sigma| = 0$), 也就是说 Σ^{-1} 不存在, 没办法拟合出多元高斯分布了, 确切的说是我们估计不出来 Σ 。

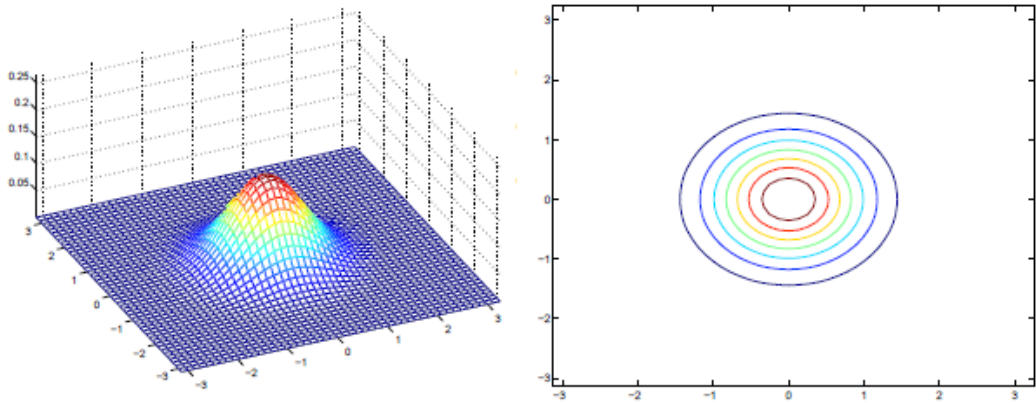
如果我们仍然想用多元高斯分布来估计样本, 那怎么办呢?

2 限制协方差矩阵

当没有足够的去估计 Σ 时, 那么只能对模型参数进行一定假设, 之前我们想估计出完全的 Σ (矩阵中的全部元素), 现在我们假设 Σ 就是对角阵 (各特征间相互独立), 那么我们只需要计算每个特征的方差即可, 最后的 Σ 只有对角线上的元素不为 0

$$\Sigma_{jj} = \frac{1}{m} \sum_{i=1}^m (x_j^{(i)} - \mu_j)^2$$

回想我们之前讨论过的二维多元高斯分布的几何特性, 在平面上的投影是个椭圆, 中心点由 μ 决定, 椭圆的形状由 Σ 决定。 Σ 如果变成对角阵, 就意味着椭圆的两个轴都和坐标轴平行了。



如果我们想对 Σ 进一步限制的话，可以假设对角线上的元素都是等值的。

$$\Sigma = \sigma^2 I$$

其中

$$\sigma^2 = \frac{1}{mn} \sum_{j=1}^n \sum_{i=1}^m (x_j^{(i)} - \mu_j)^2$$

也就是上一步对角线上元素的均值，反映到二维高斯分布图上就是椭圆变成圆。

当我们要估计出完整的 Σ 时，我们需要 $m > n+1$ 才能保证在最大似然估计下得出的 Σ 是非奇异的。然而在上面的任何一种假设限定条件下，只要 $m >= 2$ 都可以估计出限定的 Σ 。

这样做的缺点也是显而易见的，我们认为特征间独立，这个假设太强。接下来，我们给出一种称为因子分析的方法，使用更多的参数来分析特征间的关系，并且不需要计算一个完整的 Σ 。

3 边缘和条件高斯分布

在讨论因子分析之前，先看看多元高斯分布中，条件和边缘高斯分布的求法。这个在后面因子分析的 EM 推导中 useful。

假设 x 是有两个随机向量组成（可以看作是将之前的 $x^{(i)}$ 分成了两部分）

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

其中 $x_1 \in \mathbb{R}^r$, $x_2 \in \mathbb{R}^s$, 那么 $x \in \mathbb{R}^{r+s}$ 。假设 x 服从多元高斯分布 $x \sim N(\mu, \Sigma)$, 其中

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

其中 $\mu_1 \in \mathbb{R}^r$, $\mu_2 \in \mathbb{R}^s$, 那么 $\Sigma_{11} \in \mathbb{R}^{r \times r}$, $\Sigma_{12} \in \mathbb{R}^{r \times s}$, 由于协方差矩阵是对称阵，因此 $\Sigma_{12} = \Sigma_{21}^T$ 。

整体看来 x_1 和 x_2 联合分布符合多元高斯分布。

那么只知道联合分布的情况下，如何求得 x_1 的边缘分布呢？从上面的 μ 和 Σ 可以看出， $E[x_1] = \mu_1$, $\text{Cov}(x_1) = E[(x_1 - \mu_1)(x_1 - \mu_1)^T] = \Sigma_{11}$ ，下面我们验证第二个结果

$$\begin{aligned} \text{Cov}(x) &= \Sigma \\ &= \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \\ &= E[(x - \mu)(x - \mu)^T] \\ &= E\left[\begin{pmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \end{pmatrix} \begin{pmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \end{pmatrix}^T\right] \\ &= E\begin{bmatrix} (x_1 - \mu_1)(x_1 - \mu_1)^T & (x_1 - \mu_1)(x_2 - \mu_2)^T \\ (x_2 - \mu_2)(x_1 - \mu_1)^T & (x_2 - \mu_2)(x_2 - \mu_2)^T \end{bmatrix}. \end{aligned}$$

由此可见，多元高斯分布的边缘分布仍然是多元高斯分布。也就是说 $x_1 \sim N(\mu_1, \Sigma_{11})$ 。

上面 $\text{Cov}(x)$ 里面有趣的是 Σ_{12} ，这个与之前计算协方差的效果不同。之前的协方差矩阵都是针对一个随机变量（多维向量）来说的，而 Σ_{12} 评价的是两个随机向量之间的关系。比如 $x_1 = \{\text{身高}, \text{体重}\}$, $x_2 = \{\text{性别}, \text{收入}\}$ ，那么 Σ_{11} 求的是身高与身高，身高与体重，体重与体重的协方差。而 Σ_{12} 求的是身高与性别，身高与收入，体重与性别，体重与收入的协方差，看起来与之前的大不一样，比较诡异的求法。

上面求的是边缘分布，让我们考虑一下条件分布的问题，也就是 $x_1|x_2$ 的问题。根据多元高斯分布的定义， $x_1|x_2 \sim N(\mu_{1|2}, \Sigma_{1|2})$ 。

且

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

$$\Sigma_{1|2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}. \quad (2)$$

这是我们接下来计算时需要的公式，这两个公式直接给出，没有推导过程。如果想了解具体的推导过程，可以参见 Chuong B. Do 写的《Gaussian processes》。

4 因子分析例子

下面通过一个简单例子，来引出因子分析背后的思想。

因子分析的实质是认为 m 个 n 维特征的训练样例 $x^{(i)}(x_1^{(i)}, x_2^{(i)}, \dots, x_n^{(i)})$ 的产生过程如下：

1、首先在一个 k 维的空间中按照多元高斯分布生成 m 个 $z^{(i)}$ (k 维向量)，即

$$z^{(i)} \sim N(0, I)$$

2、然后存在一个变换矩阵 $\Lambda \in \mathbb{R}^{n \times k}$ ，将 $z^{(i)}$ 映射到 n 维空间中，即

$$\Lambda z^{(i)}$$

因为 $z^{(i)}$ 的均值是 0 ，映射后仍然是 0 。

3、然后将 $\Lambda z^{(i)}$ 加上一个均值 μ (n 维), 即

$$\mu + \Lambda z^{(i)}$$

对应的意义是将变换后的 $\Lambda z^{(i)}$ (n 维向量) 移动到样本 $x^{(i)}$ 的中心点 μ 。

4、由于真实样例 $x^{(i)}$ 与上述模型生成的有误差, 因此我们继续加上误差 ϵ (n 维向量), 而且 ϵ 符合多元高斯分布, 即

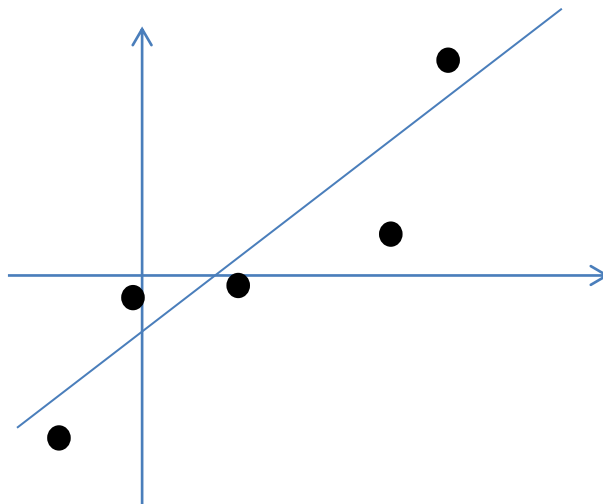
$$\epsilon \sim N(0, \Psi)$$
$$\mu + \Lambda z^{(i)} + \epsilon$$

5、最后的结果认为是真实的训练样例 $x^{(i)}$ 的生成公式

$$x^{(i)} = \mu + \Lambda z^{(i)} + \epsilon$$

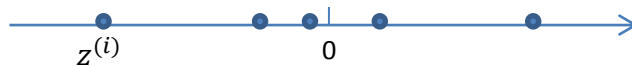
让我们使用一种直观方法来解释上述过程:

假设我们有 $m=5$ 个 2 维的样本点 $x^{(i)}$ (两个特征), 如下:



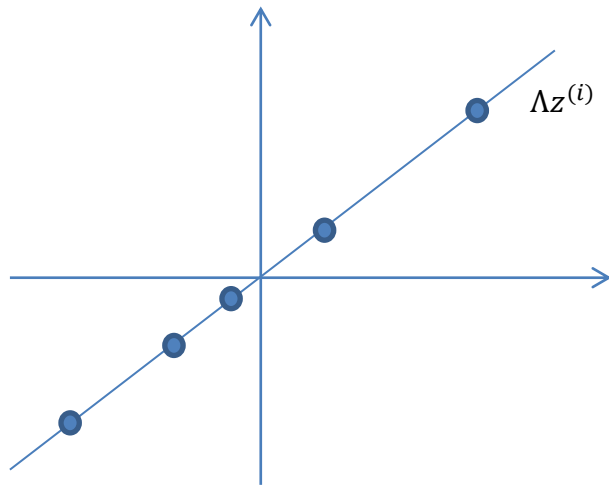
那么按照因子分析的理解, 样本点的生成过程如下:

1、我们首先认为在 1 维空间 (这里 $k=1$), 存在着按正态分布生成的 m 个点 $z^{(i)}$, 如下

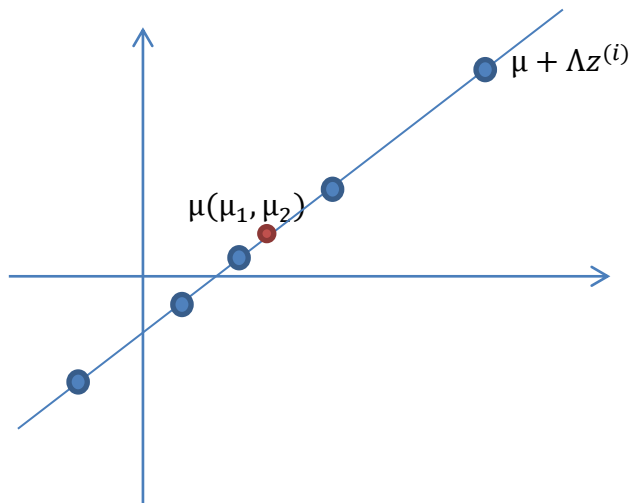


均值为 0, 方差为 1。

2、然后使用某个 $\Lambda = (a, b)^T$ 将一维的 z 映射到 2 维, 图形表示如下:

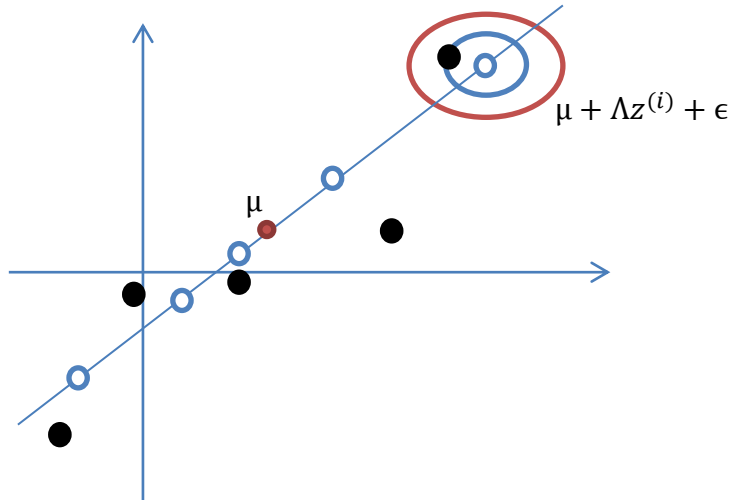


- 3、之后加上 $\mu(\mu_1, \mu_2)^T$ ，即将所有点的横坐标移动 μ_1 ，纵坐标移动 μ_2 ，将直线移到一个位置，使得直线过点 μ ，原始左边轴的原点现在为 μ （红色点）。



然而，样本点不可能这么规则，在模型上会有一些偏差，因此我们需要将上步生成的点做一些扰动（误差），扰动 $\epsilon \sim N(0, \Psi)$ 。

- 4、加入扰动后，我们得到黑色样本 $x^{(i)}$ 如下：



5、其中由于 z 和 ϵ 的均值都为 0，因此 μ 也是原始样本点（黑色点）的均值。

由以上的直观分析，我们知道了因子分析其实就是认为高维样本点实际上是由低维样本点经过高斯分布、线性变换、误差扰动生成的，因此高维数据可以使用低维来表示。

5 因子分析模型

上面的过程是从隐含随机变量 z 经过变换和误差扰动来得到观测到的样本点。其中 z 被称为因子，是低维的。

我们将式子再列一遍如下：

$$z \sim N(0, I)$$

$$\epsilon \sim N(0, \Psi)$$

$$x = \mu + \Lambda z + \epsilon$$

其中误差 ϵ 和 z 是独立的。

下面使用的因子分析表示方法是矩阵表示法，在参考资料中给出了一些其他的表示方法，如果不明白矩阵表示法，可以参考其他资料。

矩阵表示法认为 z 和 x 联合符合多元高斯分布，如下

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

求 μ_{zx} 之前需要求 $E[x]$

$$\begin{aligned} E[x] &= E[\mu + \Lambda z + \epsilon] \\ &= \mu + \Lambda E[z] + E[\epsilon] \end{aligned}$$

$$= \mu$$

我们已知 $E[z]=0$ ，因此

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

下一步是计算 Σ ，

其中 $\Sigma_{zz} = Cov(z) = I$

接着求 Σ_{zx}

$$\begin{aligned} E[(z - E[z])(x - E[x])^T] &= E[z(\mu + \Lambda z + \epsilon - \mu)^T] \\ &= E[zz^T]\Lambda^T + E[z\epsilon^T] \\ &= \Lambda^T. \end{aligned}$$

这个过程中利用了 z 和 ϵ 独立假设 ($E[z\epsilon^T] = E[z]E[\epsilon^T] = 0$)。并将 Λ 看作已知变量。

接着求 Σ_{xx}

$$\begin{aligned} E[(x - E[x])(x - E[x])^T] &= E[(\mu + \Lambda z + \epsilon - \mu)(\mu + \Lambda z + \epsilon - \mu)^T] \\ &= E[\Lambda zz^T \Lambda^T + \epsilon z^T \Lambda^T + \Lambda z \epsilon^T + \epsilon \epsilon^T] \\ &= \Lambda E[zz^T] \Lambda^T + E[\epsilon \epsilon^T] \\ &= \Lambda \Lambda^T + \Psi. \end{aligned}$$

然后得出联合分布的最终形式

$$\begin{bmatrix} z \\ x \end{bmatrix} \sim \mathcal{N} \left(\begin{bmatrix} \vec{0} \\ \mu \end{bmatrix}, \begin{bmatrix} I & \Lambda^T \\ \Lambda & \Lambda \Lambda^T + \Psi \end{bmatrix} \right).$$

从上式中可以看出 x 的边缘分布 $x \sim N(\mu, \Lambda \Lambda^T + \Psi)$

那么对样本 $\{x^{(i)}; i = 1, \dots, m\}$ 进行最大似然估计

$$\ell(\mu, \Lambda, \Psi) = \log \prod_{i=1}^m \frac{1}{(2\pi)^{n/2} |\Lambda \Lambda^T + \Psi|} \exp \left(-\frac{1}{2} (x^{(i)} - \mu)^T (\Lambda \Lambda^T + \Psi)^{-1} (x^{(i)} - \mu) \right).$$

然后对各个参数求偏导数不就得到各个参数的值了么？

可惜我们得不到 **closed-form**。想想也是，如果能得到，还干嘛将 z 和 x 放在一起求联合分布呢。根据之前对参数估计的理解，在有隐含变量 z 时，我们可以考虑使用 **EM** 来进行估计。

6 因子分析的 EM 估计

我们先来明确一下各个参数， z 是隐含变量， μ, Λ, Ψ 是待估参数。

回想 **EM** 两个步骤：

循环重复直到收敛 {

(E 步) 对于每一个 i , 计算

$$Q_i(z^{(i)}) := p(z^{(i)}|x^{(i)}; \theta)$$

(M 步) 计算

$$\theta := \arg \max_{\theta} \sum_i \sum_{z^{(i)}} Q_i(z^{(i)}) \log \frac{p(x^{(i)}, z^{(i)}; \theta)}{Q_i(z^{(i)})}.$$

我们套用一下:

(E 步):

$$Q_i(z^{(i)}) := p(z^{(i)}|x^{(i)}; \mu, \Lambda, \Psi)$$

根据第 3 节的条件分布讨论,

$$z^{(i)}|x^{(i)}; \mu, \Lambda, \Psi \sim N(\mu_{z^{(i)}|x^{(i)}}, \Sigma_{z^{(i)}|x^{(i)}})$$

因此

$$\begin{aligned} \mu_{z^{(i)}|x^{(i)}} &= \Lambda^T (\Lambda \Lambda^T + \Psi)^{-1} (x^{(i)} - \mu), \\ \Sigma_{z^{(i)}|x^{(i)}} &= I - \Lambda^T (\Lambda \Lambda^T + \Psi)^{-1} \Lambda. \end{aligned}$$

那么根据多元高斯分布公式, 得到

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

(M 步):

直接写要最大化的目标是

$$\sum_{i=1}^m \int_{z^{(i)}} Q_i(z^{(i)}) \log \frac{p(x^{(i)}, z^{(i)}; \mu, \Lambda, \Psi)}{Q_i(z^{(i)})} dz^{(i)}$$

其中待估参数是 μ, Λ, Ψ

下面我们重点求 Λ 的估计公式

首先将上式简化为:

$$\sum_{i=1}^m \int_{z^{(i)}} Q_i(z^{(i)}) [\log p(x^{(i)}|z^{(i)}; \mu, \Lambda, \Psi) + \log p(z^{(i)}) - \log Q_i(z^{(i)})] dz^{(i)} \quad (5)$$

$$= \sum_{i=1}^m E_{z^{(i)} \sim Q_i} [\log p(x^{(i)}|z^{(i)}; \mu, \Lambda, \Psi) + \log p(z^{(i)}) - \log Q_i(z^{(i)})] \quad (6)$$

这里 $z^{(i)} \sim Q_i$ 表示 $z^{(i)}$ 服从 Q_i 分布。然后去掉与 Λ 不相关的项 (后两项), 得

$$\begin{aligned}
& \sum_{i=1}^m \mathbb{E} [\log p(x^{(i)}|z^{(i)}; \mu, \Lambda, \Psi)] \\
&= \sum_{i=1}^m \mathbb{E} \left[\log \frac{1}{(2\pi)^{n/2} |\Psi|^{1/2}} \exp \left(-\frac{1}{2} (x^{(i)} - \mu - \Lambda z^{(i)})^T \Psi^{-1} (x^{(i)} - \mu - \Lambda z^{(i)}) \right) \right] \\
&= \sum_{i=1}^m \mathbb{E} \left[-\frac{1}{2} \log |\Psi| - \frac{n}{2} \log(2\pi) - \frac{1}{2} (x^{(i)} - \mu - \Lambda z^{(i)})^T \Psi^{-1} (x^{(i)} - \mu - \Lambda z^{(i)}) \right]
\end{aligned}$$

去掉不相关的前两项后，对 Λ 进行导，

$$\begin{aligned}
& \nabla_{\Lambda} \sum_{i=1}^m -\mathbb{E} \left[\frac{1}{2} (x^{(i)} - \mu - \Lambda z^{(i)})^T \Psi^{-1} (x^{(i)} - \mu - \Lambda z^{(i)}) \right] \\
&= \sum_{i=1}^m \nabla_{\Lambda} \mathbb{E} \left[-\text{tr} \frac{1}{2} z^{(i)T} \Lambda^T \Psi^{-1} \Lambda z^{(i)} + \text{tr} z^{(i)T} \Lambda^T \Psi^{-1} (x^{(i)} - \mu) \right] \\
&= \sum_{i=1}^m \nabla_{\Lambda} \mathbb{E} \left[-\text{tr} \frac{1}{2} \Lambda^T \Psi^{-1} \Lambda z^{(i)} z^{(i)T} + \text{tr} \Lambda^T \Psi^{-1} (x^{(i)} - \mu) z^{(i)T} \right] \\
&= \sum_{i=1}^m \mathbb{E} \left[-\Psi^{-1} \Lambda z^{(i)} z^{(i)T} + \Psi^{-1} (x^{(i)} - \mu) z^{(i)T} \right]
\end{aligned}$$

第一步到第二步利用了 $\text{tr} \mathbf{a} = \mathbf{a}$ (\mathbf{a} 是实数时) 和 $\text{tr} \mathbf{AB} = \text{tr} \mathbf{BA}$ 。最后一步利用了

$$\nabla_{\mathbf{A}} \text{tr} \mathbf{ABA}^T \mathbf{C} = \mathbf{CAB} + \mathbf{C}^T \mathbf{AB}^T$$

tr 就是求一个矩阵对角线上元素和。

最后让其值为 0，并且化简得

$$\sum_{i=1}^m \Lambda \mathbb{E}_{z^{(i)} \sim Q_i} [z^{(i)} z^{(i)T}] = \sum_{i=1}^m (x^{(i)} - \mu) \mathbb{E}_{z^{(i)} \sim Q_i} [z^{(i)T}].$$

然后得到

$$\Lambda = \left(\sum_{i=1}^m (x^{(i)} - \mu) \mathbb{E}_{z^{(i)} \sim Q_i} [z^{(i)T}] \right) \left(\sum_{i=1}^m \mathbb{E}_{z^{(i)} \sim Q_i} [z^{(i)} z^{(i)T}] \right)^{-1}. \quad (7)$$

到这里我们发现，这个公式有点眼熟，与之前回归中的最小二乘法矩阵形式类似

$$“\theta^T = (y^T X)(X^T X)^{-1}.”$$

这里解释一下两者的相似性，我们这里的 \mathbf{x} 是 \mathbf{z} 的线性函数（包含了一定的噪声）。在 \mathbf{E} 步得到 \mathbf{z} 的估计后，我们找寻的 Λ 实际上是 \mathbf{x} 和 \mathbf{z} 的线性关系。而最小二乘法也是去找特征和结果直接的线性关系。

到这还没完，我们需要求得括号里面的值

根据我们之前对 $z|x$ 的定义，我们知道

$$\begin{aligned} E_{z^{(i)} \sim Q_i} [z^{(i)T}] &= \mu_{z^{(i)}|x^{(i)}}^T \\ E_{z^{(i)} \sim Q_i} [z^{(i)} z^{(i)T}] &= \mu_{z^{(i)}|x^{(i)}} \mu_{z^{(i)}|x^{(i)}}^T + \Sigma_{z^{(i)}|x^{(i)}}. \end{aligned}$$

第一步根据 z 的条件分布得到，第二步根据 $\text{Cov}(Y) = E[YY^T] - E[Y]E[Y]^T$ 得到将上面的结果代入 (7) 中得到

$$\Lambda = \left(\sum_{i=1}^m (x^{(i)} - \mu) \mu_{z^{(i)}|x^{(i)}}^T \right) \left(\sum_{i=1}^m \mu_{z^{(i)}|x^{(i)}} \mu_{z^{(i)}|x^{(i)}}^T + \Sigma_{z^{(i)}|x^{(i)}} \right)^{-1}. \quad (8)$$

至此，我们得到了 Λ ，注意一点是 $E[z]$ 和 $E[zz^T]$ 的不同，后者需要 z 的协方差。

其他参数的迭代公式如下：

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

均值 μ 在迭代过程中值不变。

$$\Phi = \frac{1}{m} \sum_{i=1}^m x^{(i)} x^{(i)T} - x^{(i)} \mu_{z^{(i)}|x^{(i)}}^T \Lambda^T - \Lambda \mu_{z^{(i)}|x^{(i)}} x^{(i)T} + \Lambda (\mu_{z^{(i)}|x^{(i)}} \mu_{z^{(i)}|x^{(i)}}^T + \Sigma_{z^{(i)}|x^{(i)}}) \Lambda^T$$

然后将 Φ 上的对角线上元素抽取出来放到对应的 Ψ 中，就得到了 Ψ 。

7 总结

根据上面的 EM 的过程，要对样本 X 进行因子分析，只需知道要分解的因子数 (z 的维度) 即可。通过 EM，我们能够得到转换矩阵 Λ 和误差协方差 Ψ 。

因子分析实际上是降维，在得到各个参数后，可以求得 z 。但是 z 的各个参数含义需要自己去琢磨。

下面从一个 ppt 中摘抄几段话来进一步解释因子分析。

因子分析(factor analysis)是一种数据简化的技术。它通过研究众多变量之间的内部依赖关系，探求观测数据中的基本结构，并用少数几个假想变量来表示其基本的数据结构。这几个假想变量能够反映原来众多变量的主要信息。原始的变量是可观的显在变量，而假想变量是不可观测的潜在变量，称为因子。

例如，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。而这三个公共因子可以表示为：

$$x_i = \mu_i + \alpha_{i1}F_1 + \alpha_{i2}F_2 + \alpha_{i3}F_3 + \varepsilon_i \quad i = 1, \dots, 24$$

这里的 x_i 就是样例 x 的第 i 个分量, μ_i 就是 μ 的第 i 个分量, α_{ij} 就是 Λ 的第 i 行第 j 列元素, F_j 是 z 的第 j 个分量, ε_i 是 $\varepsilon^{(i)}$ 。

称 F_j 是不可观测的潜在因子。24 个变量共享这三个因子, 但是每个变量又有自己的个性, 不被包含的部分 ε_i , 称为特殊因子。

注:

因子分析与回归分析不同, 因子分析中的因子是一个比较抽象的概念, 而回归因子有非常明确的实际意义;

主成分分析与因子分析也有不同, 主成分分析仅仅是变量变换, 而因子分析需要构造因子模型。

主成分分析: 原始变量的线性组合表示新的综合变量, 即主成分;

因子分析: 潜在的假想变量和随机影响变量的线性组合表示原始变量。

PPT 地址

<http://www.math.zju.edu.cn/webpagenew/uploadfiles/attachfiles/2008123195228555.ppt>

其他值得参考的文献

An Introduction to Probabilistic Graphical Models by Jordan Chapter 14