CS28010: **Homework 3**

Li Hao

Nov. 22nd, 2017

1 Factor analysis

1.1 Linear factor analysis

We denote the observed data as x, the latent factor as y and the error as ϵ. Suppose $y \sim N(y|\mu, \Lambda)$, $\epsilon \sim N(\epsilon|0, \Sigma)$ and $x = Ay + \epsilon$, where A is an $n \times m$ matrix, n is the dimension of x, m is the dimension of y and $m < n$. Please explain why there is more than one solution that satisfy $E(xx^T) = A\Lambda A^T + \Sigma$. When Σ is not a general positive definite matrix, but a diagonal matrix, how many solution exists? And if $\Sigma = \sigma^2\mathbf{I}$, how many solution exists?

Solution

There are 4 uncertainties in factor analysis model:

- **rotation uncertainty**: since the covariance matrix Λ of y is diagonal, there is no rotation uncertainty for y.
- **scale uncertainty**: $y \sim N(y|\mu, \Lambda)$, which means $y \in \mathbb{R}^m$. Hence there are always scale uncertainty for y.
- **addition uncertainty**: when Σ is a general positive definite matrix or a diagonal matrix, there exists addition uncertainty. But there is no addition uncertainty when $\Sigma = \sigma^2\mathbf{I}$.
- **dimension uncertainty**: as the dimension of y is m, there is no dimension uncertainty.

Even if assuming $\Sigma = \sigma^2\mathbf{I}$ can cancel the **addition uncertainty**, there still exist **rotation uncertainty** and **scale uncertainty** in $A\Lambda A^T$. So there are always multiple solutions for all cases.

1.2 Binary factor analysis

If y is a latent factor where each dimension is an independent variable that subjects to a different Bernoulli distribution, what are the answers to the above three questions?

Solution

If y_i subjects to Bernoulli distribution, then for each dimension of y,

$$y_i \in \{0, 1\}$$

Hence there are no longer exists **rotation uncertainty** and **scale uncertainty**, since either rotate matrix or scale matrix will cause $y_i \notin \{0, 1\}$. Therefore,

- if $Sigma$ is not a general positive definite matrix, but a diagonal matrix, there still are **addition uncertainty**. So there are multiple solutions.
- if $\Sigma = \sigma^2\mathbf{I}$, there are no uncertainty. Hence we can get the only one solution.

$$\begin{bmatrix} \alpha_1^T \\ \alpha_2^T \\ \vdots \\ \alpha_k^T \end{bmatrix}$$

2 Projection

2.1 Orthogonal projection

Suppose we have a hyperplane whose orthogonal basis are $\alpha_1, \alpha_2, \cdots, \alpha_k, k < n$. Now we have a n-dimensional vector $\mathbf{x}$ and we want to apply an orthogonal projection on the hyperplane. Please compute the corresponding projection matrix P.

Solution

Assuming all vector is cloumn vector($v^T = [v_1, v_2, \cdots, v_n]$)

Denote $\mathbf{A} = [\alpha_1, \alpha_2, \cdots, \alpha_k]$ is a $n \times k$ matrix. Denote the orthogonally projected vector as $\hat{\mathbf{x}}$.
Since we have $\alpha_i^T(\mathbf{x} - \hat{\mathbf{x}}) = 0$, then

$$\begin{aligned} \mathbf{A}^T(\mathbf{x} - \hat{\mathbf{x}}) &= \mathbf{0} \\ \mathbf{A}^T\mathbf{x} &= \mathbf{A}^T\hat{\mathbf{x}} \end{aligned}$$

we also have that

$$\hat{\mathbf{x}} = c_1\alpha_1 + c_1\alpha_2 + \cdots + c_k\alpha_k = \mathbf{A}\begin{bmatrix} \alpha_1^T \\ \alpha_2^T \\ \vdots \\ \alpha_k^T \end{bmatrix} = \mathbf{Ac}$$

Hence,

$$\begin{aligned} \mathbf{A}^T\mathbf{x} &= \mathbf{A}^T\hat{\mathbf{x}} \\ &= \mathbf{A}^T\mathbf{Ac} \end{aligned}$$

so we get $\mathbf{c} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{x}$ Then

$$\hat{\mathbf{x}} = \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{x}$$

Therefore, the corresponding projection matrix $\mathbf{P}$ is:

$$\mathbf{P} = \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T$$

3 Clustering

3.1 Comparison between Gaussian mixture model and k-means

Please add constraints to Gaussian mixture model so that it degenerates into k-means algorithm.

Solution

The constraint is that: the covariance matrices of the K mixture components in GMM are:

$$\Sigma_1 = \Sigma_2 = \cdots = \Sigma_k = \sigma^2\mathbf{I} \text{ and } \sigma^2 \to 0$$

Then the distribution of the k-th component can be written as:

$$p(\mathbf{x}|\mu_k, \Sigma_k) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{1}{2\sigma^2}||\mathbf{x} - \mu_k||^2\right\}$$

And consider the EM algorithm for a mixture of K Gaussians of this form.

In the expectation step:

$$\gamma_{nk} = p(z_n = k|\mathbf{x}_n, \mu, \Sigma, \pi) = \frac{\pi_k \exp\{-||\mathbf{x}_n - \mu_k||^2/2\sigma\}}{\sum_i \pi_i \exp\{-||\mathbf{x}_n - \mu_i||^2/2\sigma\}}$$

if $\sigma^2 \to 0$, then

$$\gamma_{nk} = r_{nk} \to \begin{cases} 1 & \text{if } k = \text{arg min}_i ||\mathbf{x}_n - \mu_i||^2 \\ 0 & \text{otherwise} \end{cases}$$

which is a nearly hard assignments just like k-means algorithm.

In the maximization step:

$$\mathbb{E}_{\mathbf{Z}}[\ln p(\mathbf{X}, \mathbf{Z}|\mu, \Sigma, \pi)] \to -\frac{1}{2} \sum_{n=1}^{N} \sum_{k=1}^{K} \gamma_{nk} ||\mathbf{x}_n - \mu_k||^2 + const$$

which is also almost same as the maximization step in the k-means algorithm which maximize:

$$\mathbf{J} = \sum_{n=1}^{N} \sum_{k=1}^{K} r_{nk} ||\mathbf{x}_n - \mu_k||^2$$

Therefore, in the given constraint above, Gaussian mixture model degenerates into kmeans algorithm.

4 Optional summary work

Note: this is an optional homework. Please compare PCA, FA and ICA.
Solution