

# Factor Analysis Theory

May 6, 2022

## 1 Factor analysis

### 1.1 Why factor analysis?

When we have data  $x^{(i)} \in \mathbb{R}^n$  coming from a mixture of several Gaussians, we can use the EM Algorithm to fit the mixture model. However, we usually have sufficient data to discern the multiple-Gaussian structure.

Assuming we have  $m$  observations of  $n$  dimensional data. Consider a setting in which  $n \gg m$ . If we model the data as Gaussian and estimate the mean and covariance as usual maximum likelihood estimators,

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

we would find that matrix  $\Sigma$  is singular. Therefore, the Gaussian probability distribution function cannot be estimated.

More generally, unless  $m$  exceeds  $n$  by reasonable amount, the maximum likelihood estimates of the mean and covariance may be quite poor. In this case, we can use the Factor Analysis model to analyse underlying factor and reduce dimensionality of the data.

### 1.2 The Factor analysis model

In the Factor analysis model, we create a joint probability distribution on  $(x, z)$  where  $z \in \mathbb{R}^k$  is a latent random variable, as follows:

$$z \sim \mathcal{N}(0, I)$$
$$x|z \sim \mathcal{N}(\mu + \Lambda z, \Psi)$$

the parameters of our model being  $\mu \in \mathbb{R}^n$ , the matrix,  $\Lambda \in \mathbb{R}^{n \times k}$  and the diagonal matrix  $\Psi \in \mathbb{R}^{n \times n}$ .

We can also define the factor analysis model according to

$$z \sim \mathcal{N}(0, I)$$
$$\epsilon \sim \mathcal{N}(0, \Psi)$$

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

The random variables  $z$  and  $x$  have joint Gaussian distribution

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

$E[z] = 0$  because  $z \sim \mathcal{N}(0, I)$

$$\begin{aligned} E[x] &= E[\mu + \Lambda z + \epsilon] \\ &= \mu + \Lambda E[z] + E[\epsilon] \\ &= \mu \end{aligned} \tag{1}$$

Therefore,

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

Now, to find  $\Sigma$  where

$$\Sigma = \begin{bmatrix} \Sigma_{zz} & \Sigma_{zx} \\ \Sigma_{xz} & \Sigma_{xx} \end{bmatrix}$$

We calculate  $\Sigma_{zz} = E[(z - E[z])(z - E[z])^T]$ ,  $\Sigma_{zx} = E[(z - E[z])(x - E[x])^T]$ ,  $\Sigma_{xz} = E[(x - E[x])(z - E[z])^T]$  and  $\Sigma_{xx} = E[(x - E[x])(x - E[x])^T]$

After calculating all these values, we obtain:

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

We can see that the marginal distribution of  $x$  is given by  $x \sim \mathcal{N}(\mu, \Lambda \Lambda^T + \Psi)$

If given a training set  $\{x^{(i)}; i = 1, \dots, m\}$ , the log likelihood of the parameters will be

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

### 1.3 EM For Factor Analysis

#### 1.3.1 E - Step:

We need to compute

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

Substituting distribution in Equation (3) into formulas (1-2), we find that,

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

where,

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

Using the above definitions, we have

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

### 1.3.2 M - Step:

For M-step, we need to maximize:

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

with respect to  $\mu, \Lambda, \psi$

On solving for  $\Lambda$  we will get the following expression:

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

Note the close relationship between this equation and Normal Equation:

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

The analogy here is that  $x$ 's are the linear function of  $z$ 's (plus noise). Given the guesses for  $z$  from E-step we will now try to estimate the unknown linearity  $\Lambda$  relating the  $x$ 's and  $z$ 's

The M-step update for Lambda

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

Lastly, M-step optimization steps for the other parameters can be given by:

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

This can be calculated just once and needs not to be further updated.

Similarly,

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

[ ]: