

Multivariate Statistical Analysis

Lecture 16

Fudan University

luoluo@fudan.edu.cn

- 1 Factor Analysis
- 2 Probabilistic Principle Component Analysis
- 3 The Expectation-Maximization Algorithm
- 4 Course Summary

- 1 Factor Analysis
- 2 Probabilistic Principle Component Analysis
- 3 The Expectation-Maximization Algorithm
- 4 Course Summary

Let the observable vector $\mathbf{y} \in \mathbb{R}^p$ be written as

$$\mathbf{y} = \mathbf{W}\mathbf{x} + \boldsymbol{\mu} + \boldsymbol{\epsilon},$$

where

- 1 $\mathbf{W} \in \mathbb{R}^{p \times q}$ is the loading matrix (parameter),
- 2 $\mathbf{x} \in \mathbb{R}^q$ is the common factor (parameter/random vector),
- 3 $\boldsymbol{\mu} \in \mathbb{R}^p$ is the mean vector (parameter),
- 4 $\boldsymbol{\epsilon} \in \mathbb{R}^p$ is the specific factor (random vector).

The model is similar to regression, but \mathbf{x} is unobserved.

Factor Analysis

Example of sports games:

$$\mathbf{y} = \mathbf{W}\mathbf{x} + \boldsymbol{\mu} + \boldsymbol{\epsilon}.$$

- 1 \mathbf{y} : performance in real-world
- 2 \mathbf{W} : system of the game
- 3 \mathbf{x} : attributes in the game
- 4 $\boldsymbol{\mu}$: average attributes
- 5 $\boldsymbol{\epsilon}$: noise/exception

57 Games	82 Games	80 Games	48 Games
22.3 Points	17.5 Points	18.3 Points	25.0 Points
51.9% FG%	52.2% FG%	55.2% FG%	51.6% FG%
0.0% 3P%	0.0% 3P%	0.0% 3P%	0.0% 3P%
85.3% FT%	80.9% FT%	86.2% FT%	86.2% FT%
10.2 Rebounds	9.4 Rebounds	9.4 Rebounds	9.4 Rebounds
1.5 Assists	2.0 Assists	2.0 Assists	2.0 Assists
0.5 Steals	0.4 Steals	0.4 Steals	0.4 Steals
1.6 Blocks	2.0 Blocks	2.0 Blocks	2.0 Blocks
2.6 Turnovers	2.5 Turnovers	3.5 Turnovers	3.5 Turnovers
25.6 PER	21.9 PER	26.5 PER	26.5 PER
0.211 ws/48	0.202 ws/48	0.220 ws/48	0.220 ws/48



- 1 Factor Analysis
- 2 Probabilistic Principle Component Analysis**
- 3 The Expectation-Maximization Algorithm
- 4 Course Summary

Probabilistic Principle Component Analysis

Let $\mathbf{y}_1, \dots, \mathbf{y}_N \in \mathbb{R}^p$ be N independent observations and we have

$$\mathbf{y}_\alpha = \mathbf{W}\mathbf{x}_\alpha + \boldsymbol{\mu} + \epsilon_\alpha,$$

where

$$\mathbf{x}_\alpha \sim \mathcal{N}_q(\mathbf{0}, \mathbf{I}) \quad \text{and} \quad \epsilon_\alpha \sim \mathcal{N}_p(\mathbf{0}, \sigma^2 \mathbf{I})$$

are independent for some $\sigma^2 > 0$ and $q < \min\{N, p\}$.

We target to estimate parameters

$$\mathbf{W} \in \mathbb{R}^{p \times q}, \quad \boldsymbol{\mu} \in \mathbb{R}^p \quad \text{and} \quad \sigma \in (0, +\infty)$$

by maximum likelihood estimation for given $\mathbf{y}_1, \dots, \mathbf{y}_N$.

Probabilistic Principle Component Analysis

Consider that

$$\mathbf{y}_\alpha \sim \mathcal{N}_p(\boldsymbol{\mu}, \mathbf{W}\mathbf{W}^\top + \sigma^2\mathbf{I}).$$

We construct the likelihood function

$$\begin{aligned} & L(\boldsymbol{\mu}, \mathbf{W}, \sigma^2) \\ &= \prod_{\alpha=1}^N \frac{1}{\sqrt{(2\pi)^p \det(\boldsymbol{\Sigma})}} \exp\left(-\frac{1}{2}(\mathbf{y}_\alpha - \boldsymbol{\mu})^\top (\mathbf{W}\mathbf{W}^\top + \sigma^2\mathbf{I})^{-1}(\mathbf{y}_\alpha - \boldsymbol{\mu})\right), \end{aligned}$$

then we have

$$\begin{aligned} & \ln L(\boldsymbol{\mu}, \mathbf{W}, \sigma^2) \\ & \propto -\frac{N}{2} \ln \det(\mathbf{W}\mathbf{W}^\top + \sigma^2\mathbf{I}) - \frac{1}{2} \sum_{\alpha=1}^N (\mathbf{y}_\alpha - \boldsymbol{\mu})^\top (\mathbf{W}\mathbf{W}^\top + \sigma^2\mathbf{I})^{-1}(\mathbf{y}_\alpha - \boldsymbol{\mu}). \end{aligned}$$

The Maximum Likelihood Estimators

The maximum likelihood estimators of $\boldsymbol{\mu}$, \mathbf{W} and σ^2 are

$$\hat{\boldsymbol{\mu}} = \bar{\mathbf{y}} = \frac{1}{N} \sum_{\alpha=1}^N \mathbf{y}_{\alpha}, \quad \hat{\mathbf{W}} = \mathbf{U}_q (\boldsymbol{\Lambda}_q - \hat{\sigma}^2 \mathbf{I}) \mathbf{R} \quad \text{and} \quad \hat{\sigma}^2 = \frac{1}{p-q} \sum_{j=q+1}^p \lambda_j,$$

where

- 1 $\boldsymbol{\Lambda}_q \in \mathbb{R}^{q \times q}$ is diagonal with the largest q eigenvalues $\lambda_1, \dots, \lambda_q$ of

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{N} \sum_{\alpha=1}^N (\mathbf{y}_{\alpha} - \bar{\mathbf{y}})(\mathbf{y}_{\alpha} - \bar{\mathbf{y}})^{\top};$$

- 2 $\mathbf{U}_q \in \mathbb{R}^{p \times q}$ is orthogonal column consisting of the eigenvectors associate with $\lambda_1, \dots, \lambda_q$;
- 3 $\mathbf{R} \in \mathbb{R}^{q \times q}$ is any orthogonal matrix.

The Maximum Likelihood Estimators

The maximum likelihood estimators also minimize the error with respect to Frobenius norm

$$(\hat{\mathbf{W}}, \hat{\sigma}^2) = \arg \min_{\mathbf{W} \in \mathbb{R}^{p \times q}, \sigma^2 \in \mathbb{R}^+} \left\| \hat{\mathbf{\Sigma}} - (\mathbf{W}\mathbf{W}^T + \sigma^2 \mathbf{I}) \right\|_F.$$

Outline

- 1 Factor Analysis
- 2 Probabilistic Principle Component Analysis
- 3 The Expectation-Maximization Algorithm**
- 4 Course Summary

The Expectation-Maximization Algorithm

For the model

$$\mathbf{y} = \mathbf{W}\mathbf{x} + \boldsymbol{\mu} + \epsilon,$$

where $\mathbf{x} \sim \mathcal{N}_q(\mathbf{0}, \mathbf{I})$ and $\epsilon \sim \mathcal{N}_p(\mathbf{0}, \sigma^2 \mathbf{I})$ are independent.

We regard $\{\mathbf{x}_\alpha\}_{\alpha=1}^N$ as missing data and $\{\mathbf{x}_\alpha, \mathbf{y}_\alpha\}_{\alpha=1}^N$ as the complete data, then we can achieve

$$\mathbf{y}_\alpha | \mathbf{x}_\alpha \sim \mathcal{N}_p(\mathbf{W}\mathbf{x}_\alpha + \boldsymbol{\mu}, \sigma^2 \mathbf{I})$$

and

$$\mathbf{x}_\alpha | \mathbf{y}_\alpha \sim \mathcal{N}_q(\mathbf{M}^{-1} \mathbf{W}^\top (\mathbf{y}_\alpha - \boldsymbol{\mu}), \sigma^2 \mathbf{M}^{-1}),$$

where $\mathbf{M} = \mathbf{W}^\top \mathbf{W} + \sigma^2 \mathbf{I}$.

The Expectation-Maximization Algorithm

The update of the EM algorithm

- 1 In E-step, we take the expectation

$$l_C = \mathbb{E} \left[\ln \left(\prod_{\alpha=1}^N f(\mathbf{x}_\alpha | \mathbf{y}_\alpha) \right) \right].$$

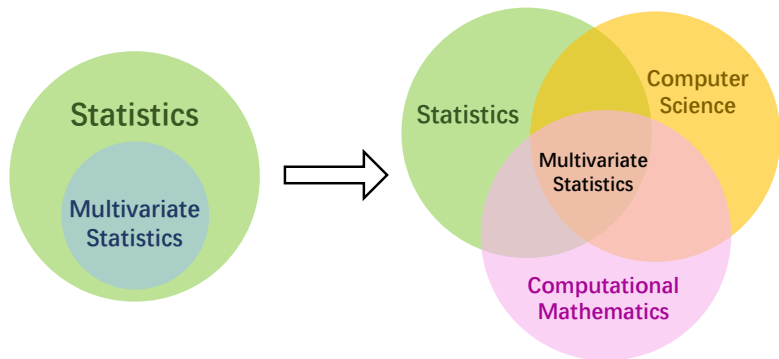
- 2 In the M-step, we maximized l_C with respect to \mathbf{W} and σ^2 :

$$\begin{aligned} \mathbf{W}_+ &= \hat{\Sigma} \mathbf{W} (\sigma^2 \mathbf{I} + \mathbf{M}^{-1} \mathbf{W}^\top \hat{\Sigma} \mathbf{W})^{-1}, \\ \sigma_+^2 &= \frac{1}{p} \text{tr} \left(\hat{\Sigma} - \hat{\Sigma} \mathbf{W} \mathbf{M}^{-1} \mathbf{W}_+^\top \right). \end{aligned}$$

Note that the computational complexity of EM is $\mathcal{O}(Npq)$, while the spectral decomposition in MLE requires $\mathcal{O}(Np^2 + p^3)$.

- 1 Factor Analysis
- 2 Probabilistic Principle Component Analysis
- 3 The Expectation-Maximization Algorithm
- 4 Course Summary**

Multivariate Statistics



Good Luck on Finals!

Final Exam



Multivariate Statistical Analysis
(DATA 13004)

Matrix Calculus



Linear Algebra

Machine Learning



Statistics



Optimization

