# Multivariate Statistical Analysis

Lecture 16

Fudan University

luoluo@fudan.edu.cn

# Outline

# Outline

# Factor Analysis

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

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

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. $\epsilon \in \mathbb{R}^p$ is the specific factor (random vector).

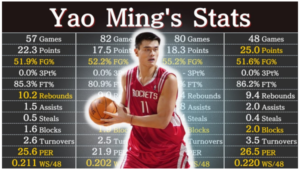
The model is similar to regression, but $\mathbf{x}$ is unobserved.

# Factor Analysis

Example of sports games:

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

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. $\epsilon$: noise/exception



# Outline

# 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

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

are independent for some $\sigma^2 > 0$ and $q < p$.

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 $\sigma^2$ are

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

where

1. $\mathbf{U}_q \in \mathbb{R}^{p\times q}$ is orthogonal column consisting of principal eigenvectors 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. $\boldsymbol{\Lambda}_q \in \mathbb{R}^{q\times q}$ is diagonal with corresponding eigenvalues $\lambda_1, \dots, \lambda_q$,
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

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

# Outline

1. Factor Analysis

2. Probabilistic Principle Component Analysis

3. The Expectation-Maximization Algorithm

# 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

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

and

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

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

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

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

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

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