Lecture 3

Multivariate normal

$$\begin{pmatrix} X_1 \\ \vdots \\ X_p \end{pmatrix} = \underset{\sim}{X} \sim N_p(\underset{\text{mean vector}}{\underset{\sim}{\mu}}, \underset{\text{covariance matrix}}{C})$$

$$f(\underset{\sim}{x}) = \frac{1}{(2\pi)^{p/2}|C|^{1/2}} \exp\left[-\frac{1}{2}(\underset{\sim}{x} - \underset{\sim}{\mu})^T C^{-1}(\underset{\sim}{x} - \underset{\sim}{\mu})\right]$$

Graphical Model

$$K = C^{-1} = \text{concentration matrix} = \begin{pmatrix} k_{11} & \cdots & k_{1p} \\ \vdots & \ddots & \vdots \\ k_{p1} & \cdots & k_{pp} \end{pmatrix}$$

$k_{ij} = 0 \Rightarrow$ no connection/edge

$k_{ij} \neq 0 \Rightarrow$ edge between i & j

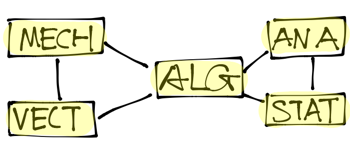
Example: Exam marks in mechanics, vectors, algebra, analysis, statistics

- standardized vars $\longrightarrow$ look at correlation matrix
- assume multivariate normality!

$$\hat{R} = \begin{pmatrix} 1 & 0.55 & 0.55 & 0.41 & 0.39 \\ & 1 & 0.61 & 0.49 & 0.44 \\ & & 1 & 0.71 & 0.66 \\ & & & 1 & 0.61 \\ & & & & 1 \end{pmatrix}$$

$$\hat{K} = \hat{R}^{-1} = \begin{array}{ccccc} \text{MECH} & \text{VECT} & \text{ALG} & \text{ANA} & \text{STAT} \\ \hline 1.60 & -0.56 & -0.51 & \cancel{0.00} & -\cancel{0.04} \\ & 1.80 & -0.66 & -\cancel{0.15} & -\cancel{0.04} \\ & & 3.04 & -1.11 & -0.86 \\ & & & 2.18 & -0.52 \\ & & & & 1.92 \end{array}$$

Graphical Model (from simple analysis)



(ANA, STAT) & (MECH, VECT) are conditionally independent given ALG.

Maximum Likelihood estimation

$\underset{\sim}{X}_1, \underset{\sim}{X}_2, \underset{\sim}{X}_3, \cdots, \underset{\sim}{X}_n$ indep $N_p(\underset{\sim}{\mu}, C) \longleftarrow$ positive definite

$$L(\underset{\sim}{\mu}, C) = \prod_{i=1}^n \left\{ \frac{1}{(2\pi)^{p/2}|C|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(\underset{\sim}{X}_i - \underset{\sim}{\mu})^T C^{-1}(\underset{\sim}{X}_i - \underset{\sim}{\mu})\right\}\right\}$$

- maximize L over $\underset{\sim}{\mu}$ & pos. def. C.

$$\ln L(\underset{\sim}{\mu}, C) = -\frac{n}{2}\ln(|C|) - \frac{1}{2}\overbrace{\sum_{i=1}^n (\underset{\sim}{X}_i - \underset{\sim}{\mu})^T C^{-1}(\underset{\sim}{X}_i - \underset{\sim}{\mu})}^{tr(C^{-1}\sum(\underset{\sim}{X}_i - \underset{\sim}{\mu})(\underset{\sim}{X}_i - \underset{\sim}{\mu})^T)} + \text{OTHER STUFF (constant)}$$

$tr(BA) = tr(AB)$

$$= -\frac{n}{2}[\ln(|C|) + \text{trace}(\underbrace{C^{-1}}_{= K}\hat{C}) + (\underset{\sim}{\mu} - \bar{\underset{\sim}{X}})^T C^{-1}(\underset{\sim}{\mu} - \bar{\underset{\sim}{X}})] \quad \text{where } \hat{C} = \frac{1}{n}\sum_{i=1}^n(\underset{\sim}{X}_i - \bar{\underset{\sim}{X}})(\underset{\sim}{X}_i - \bar{\underset{\sim}{X}})^T$$

$\Rightarrow$ MLE of $\underset{\sim}{\mu}$ is $\bar{\underset{\sim}{X}}$

$$\ln L(\cancel{\mu}, K) = -\frac{n}{2}[\text{trace}(K\hat{C}) - \ln(|K|)]$$

can be taken out

When $p < n$, then $\hat{C}$ is the MLE of C

$p \geq n$, MLE of C does not exist

Relationship to graphical models

- if assume a particular model, i.e. certain off-diagonal elements of K are equal to 0, then can maximize likelihood subject to these constraints.

Alternative approach: Graphical lasso

- assume K is "sparse" i.e. many 0 elements

maximize $\ln(|K|) - \operatorname{trace}(K\hat{C}) - \lambda \|K\|_1$

λ → non-negative tuning parameter

Where $\|K\|_1 = \sum_{i=1}^{p}\sum_{j=1}^{p} |k_{ij}|$ (R package: glasso)

or $\|K\|_1 = \sum\sum_{i \neq j} |k_{ij}|$ → summing over off-diagonal elements

Back to exam data:

- start with graphical lasso (using $\hat{C} = \hat{R}$)

$\lambda = 0.45$

$$\hat{K}_\lambda = \begin{pmatrix} 1.02 & -0.09 & -0.08 & 0 & 0 \\ & 1.03 & -0.16 & 0 & 0 \\ & & 1.14 & -0.26 & -0.20 \\ & & & 1.09 & -0.12 \\ & & & & 1.06 \end{pmatrix}$$

$$\hat{R}_\lambda = \hat{K}_\lambda^{-1} = \begin{pmatrix} 1 & \overset{=0.55}{0.10} & \overset{=0.55}{0.10} & 0.03 & 0.02 \\ & 1 & 0.16 & 0.04 & 0.03 \\ & & 1 & 0.26 & 0.21 \\ & & & 1 & 0.16 \\ & & & & 1 \end{pmatrix}$$

- correlation shrunken significantly towards 0

Now compute MLE of K assuming $(1,4), (1,5), (2,4), (2,5)$ elements are 0

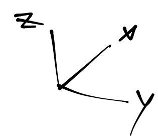
$$\hat{K} = \begin{pmatrix} 1.6 & -0.56 & -0.53 & 0 & 0 \\ & 1.79 & -0.78 & 0 & 0 \\ & & 3.22 & -1.19 & -.90 \\ & & & 2.16 & -0.52 \\ & & & & 1.92 \end{pmatrix}$$

Multivariate visualization

Data: $x_1, x_2, \ldots, x_i = \begin{pmatrix} x_{i1} \\ \vdots \\ x_{ip} \end{pmatrix}$

Q: How do we look at data if $p > 2$?

- for $p=2$, scatterplot of $\{x_{i1}\}$ vs $\{x_{i2}\}$ contains all appropriate information.
- $p=3$? Can gain illusion of 3 dimensions by using notion for 2-dim plots



Different approaches

① Find "interesting projections" of the data

meaning?

② "Grand Tour" methods
- do an exhaustive search of all 2D projections
- scatterplot movie