STAT 3690 Lecture 30

zhiyanggeezhou.github.io

Zhiyang Zhou (zhiyang.zhou@umanitoba.ca)

Apr 11, 2022

Testing the uncorrelatedness of canonical variates

- LRT for $H_0 : \boldsymbol{\Sigma}_{\mathbf{YX}} = 0$ vs. H_1 : otherwise
  - LRT statistic $\lambda = \prod_{k=1}^p (1 - \hat\rho_k^2)^{n/2}$
    - $\hat\rho_k$: the kth sample canonical correlation
    - Under H_0, $-2\ln\lambda = -n\sum_{k=1}^p \ln(1-\hat\rho_k^2) \approx \chi^2(pq)$

- Sequential inference
  - Determining r, the number of pairs of canonical variates to retain
  - Note that $\boldsymbol{\Sigma}_{\mathbf{YX}} = 0 \Leftrightarrow \rho_1 = \cdots = \rho_p = 0 \Leftrightarrow \rho_1 = 0$
    - Since $\rho_1 \ge \cdots \ge \rho_p$
  - Consider a sequence of p pairs of hypotheses: $H_{0,k} : \rho_{k-1} > 0, \rho_k = 0$ vs. $H_{1,k} : \rho_k > 0$
    - LRT statistic $\lambda_k = \prod_{\ell=k}^p (1-\hat\rho_\ell^2)^{n/2}$
      - Under $H_{0,k}$, $-2\ln\lambda_k = -n\sum_{\ell=k}^p \ln(1-\hat\rho_\ell^2) \approx \chi^2((p-k+1)(q-k+1))$
  - Stopping rules
    - p_k: the p-value associated with the testing on $H_{0,k}$ vs. $H_{1,k}$
    - $p_{(k)}$: the kth smallest value among $\{p_1, \ldots, p_p\}$
    - Holm-Bonferroni procedure (Holm (1979), Scandinavian Journal of Statistics, 6, 65–70): if $p_{(k)} < \alpha/(p+1-k)$, reject $H_{0,(k)}$ and proceed to larger p-values; otherwise EXIT.
    - B-H procedure (Benjamini & Hochberg (1995), Journal of the Royal Statistical Society, Series B., 57, 289–300):
      1. For a given level α, find $k^* = \max\{k \in \{1, \ldots, p\} \mid p_{(k)} \le k\alpha/p\}$
      2. Reject $H_{0,(k)}$ for $k = 1, \ldots, k^*$

Summary of CCA

- Dimension reduction method
  - Maximize correlation
  - Treat $\mathbf{Y}$ and $\mathbf{X}$ equally/reduce the dimension of both $\mathbf{Y}$ and $\mathbf{X}$ simultaneously
- Limitation: in need of invertible $\boldsymbol{\Sigma}_{\mathbf{Y}}$ and $\boldsymbol{\Sigma}_{\mathbf{X}}$