STAT 3690 Lecture 28

zhiyanggeezhou.github.io

Zhiyang Zhou (zhiyang.zhou@umanitoba.ca)

Apr 06, 2022

Canonical correlation analysis (CCA)

- Dimension reduction method
  - Simultaneously reducing the dimension of two random vectors $\mathbf{Y}$ and $\mathbf{X}$
  - Dropping info that has little impact on the association between $\mathbf{Y}$ and $\mathbf{X}$
- Population version
  - Random p-vector $\mathbf{Y}$ and random q-vector $\mathbf{X}$
    * $p \leq q$
    * $\mathbf{\Sigma_Y} = \mathrm{var}(\mathbf{Y}) > 0$, $\mathbf{\Sigma_X} = \mathrm{var}(\mathbf{X}) > 0$ and $\mathbf{\Sigma_{YX}} = \mathbf{\Sigma_{XY}^\top} = \mathrm{cov}(\mathbf{Y}, \mathbf{X})$
  - Vocabulary
    * (The kth pair of) canonical directions: $(\boldsymbol{a}_k \in \mathbb{R}^p, \boldsymbol{b}_k \in \mathbb{R}^q)$
    * (The kth pair of) canonical variates: $(\boldsymbol{a}_k^\top \mathbf{Y}, \boldsymbol{b}_k^\top \mathbf{X})$
    * (The kth) canonical correlation: $\rho_k = \mathrm{corr}(\boldsymbol{a}_k^\top \mathbf{Y}, \boldsymbol{b}_k^\top \mathbf{X})$
  - Goal: find $\boldsymbol{a}_k$ and $\boldsymbol{b}_k$, $k = 1, \ldots, r \leq p$, to maximize

$$\rho_k = \mathrm{corr}(\boldsymbol{a}_k^\top \mathbf{Y}, \boldsymbol{b}_k^\top \mathbf{X}) = \frac{\boldsymbol{a}_k^\top \mathbf{\Sigma_{YX}} \boldsymbol{b}_k}{\sqrt{\boldsymbol{a}_k^\top \mathbf{\Sigma_Y} \boldsymbol{a}_k}\sqrt{\boldsymbol{b}_k^\top \mathbf{\Sigma_X} \boldsymbol{b}_k}}$$

    subject to
    * $\mathrm{var}(\boldsymbol{a}_k^\top \mathbf{Y}, \boldsymbol{a}_k^\top \mathbf{Y}) = \boldsymbol{a}_k^\top \mathbf{\Sigma_Y} \boldsymbol{a}_k = 1$
    * $\mathrm{var}(\boldsymbol{b}_k^\top \mathbf{X}, \boldsymbol{b}_k^\top \mathbf{X}) = \boldsymbol{b}_k^\top \mathbf{\Sigma_X} \boldsymbol{b}_k = 1$
    * $\mathrm{cov}(\boldsymbol{a}_k^\top \mathbf{Y}, \boldsymbol{a}_\ell^\top \mathbf{Y}) = \boldsymbol{a}_k^\top \mathbf{\Sigma_Y} \boldsymbol{a}_\ell = 0$, $\ell = 1, \ldots, k-1$
    * $\mathrm{cov}(\boldsymbol{a}_k^\top \mathbf{Y}, \boldsymbol{b}_\ell^\top \mathbf{X}) = \boldsymbol{a}_k^\top \mathbf{\Sigma_{YX}} \boldsymbol{b}_\ell = 0$, $\ell = 1, \ldots, k-1$
    * $\mathrm{cov}(\boldsymbol{b}_k^\top \mathbf{X}, \boldsymbol{b}_\ell^\top \mathbf{X}) = \boldsymbol{b}_k^\top \mathbf{\Sigma_X} \boldsymbol{b}_\ell = 0$, $\ell = 1, \ldots, k-1$
    * $\mathrm{cov}(\boldsymbol{b}_k^\top \mathbf{X}, \boldsymbol{a}_\ell^\top \mathbf{Y}) = \boldsymbol{b}_k^\top \mathbf{\Sigma_{XY}} \boldsymbol{a}_\ell = 0$, $\ell = 1, \ldots, k-1$
  - Solution: Let $\mathbf{M} = \mathbf{\Sigma_Y}^{-1/2} \mathbf{\Sigma_{YX}} \mathbf{\Sigma_X}^{-1/2}$
    * $\rho_k = \sqrt{\lambda_k}$ is the kth largest singular value of $\mathbf{M}$
      · λ_k: the kth largest eigenvalue of $\mathbf{M}\mathbf{M}^\top$ (or $\mathbf{M}^\top\mathbf{M}$)
    * $\boldsymbol{a}_k = \mathbf{\Sigma_Y}^{-1/2} \boldsymbol{e}_k$
      · $\boldsymbol{e}_k$: the left-singular vector corresponding to the kth largest singular value of $\mathbf{M}$, i.e., the eigenvector corresponding to the kth largest eigenvalue of $\mathbf{M}\mathbf{M}^\top$
    * $\boldsymbol{b}_k = \mathbf{\Sigma_X}^{-1/2} \boldsymbol{f}_k$
      · $\boldsymbol{f}_k$: the right-singular vector corresponding to the kth largest singular value of $\mathbf{M}$, i.e., the eigenvector corresponding to the kth largest eigenvalue of $\mathbf{M}^\top\mathbf{M}$

- Sample version
  - $(\mathbf{Y}_1, \mathbf{X}_1), \ldots, (\mathbf{Y}_n, \mathbf{X}_n) \overset{\text{iid}}{\sim} (\mathbf{Y}, \mathbf{X})$

    - $\mathbf{Y}_i$ and $\mathbf{X}_i$ jointly sampled
    - $p \le q < n$
- $n \times p$ matrix $\mathbb{Y} = [\mathbf{Y}_1, \ldots, \mathbf{Y}_n]^\top$ and $n \times q$ matrix $\mathbb{X} = [\mathbf{X}_1, \ldots, \mathbf{X}_n]^\top$
- Sample covariance matrices
    - $\mathbf{S}_\mathbf{Y} = (n-1)^{-1} \sum_i (\mathbf{Y}_i - \bar{\mathbf{Y}})(\mathbf{Y}_i - \bar{\mathbf{Y}})^\top$
    - $\mathbf{S}_\mathbf{X} = (n-1)^{-1} \sum_i (\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})^\top$
    - $\mathbf{S}_{\mathbf{YX}} = \mathbf{S}_{\mathbf{XY}}^\top = (n-1)^{-1} \sum_i (\mathbf{Y}_i - \bar{\mathbf{Y}})(\mathbf{X}_i - \bar{\mathbf{X}})^\top$
- Vocabulary
    - (The kth pair of) sample canonical directions: $(\hat{\boldsymbol{a}}_k \in \mathbb{R}^p, \hat{\boldsymbol{b}}_k \in \mathbb{R}^q)$
    - (The kth pair of) sample canonical variates: $(\mathbb{Y}\hat{\boldsymbol{a}}_k, \mathbb{X}\hat{\boldsymbol{b}}_k)$
    - (The kth) canonical correlation: $\hat{\rho}_k$
- Goal: find $\hat{\boldsymbol{a}}_k$ and $\hat{\boldsymbol{b}}_k$, $k = 1, \ldots, r \le p$, to maximize

$$\hat{\rho}_k = \frac{\hat{\boldsymbol{a}}_k^\top \mathbf{S}_{\mathbf{YX}} \hat{\boldsymbol{b}}_k}{\sqrt{\hat{\boldsymbol{a}}_k^\top \mathbf{S}_\mathbf{Y} \hat{\boldsymbol{a}}_k}\sqrt{\hat{\boldsymbol{b}}_k^\top \mathbf{S}_\mathbf{X} \hat{\boldsymbol{b}}_k}}$$

  subject to
    - $\hat{\boldsymbol{a}}_k^\top \mathbf{S}_\mathbf{Y} \hat{\boldsymbol{a}}_k = 1$
    - $\hat{\boldsymbol{b}}_k^\top \mathbf{S}_\mathbf{X} \hat{\boldsymbol{b}}_k = 1$
    - $\hat{\boldsymbol{a}}_k^\top \mathbf{S} \hat{\boldsymbol{a}}_\ell = 0$, $\ell = 1, \ldots, k-1$
    - $\hat{\boldsymbol{a}}_k^\top \mathbf{S}_{\mathbf{YX}} \hat{\boldsymbol{b}}_\ell = 0$, $\ell = 1, \ldots, k-1$
    - $\hat{\boldsymbol{b}}_k^\top \mathbf{S}_\mathbf{X} \hat{\boldsymbol{b}}_\ell = 0$, $\ell = 1, \ldots, k-1$
    - $\hat{\boldsymbol{b}}_k^\top \mathbf{S}_{\mathbf{XY}} \hat{\boldsymbol{a}}_\ell = 0$, $\ell = 1, \ldots, k-1$
- Solution: Let $\widehat{\mathbf{M}} = \mathbf{S}_\mathbf{Y}^{-1/2} \mathbf{S}_{\mathbf{YX}} \mathbf{S}_\mathbf{X}^{-1/2}$
    - $\hat{\rho}_k = \sqrt{\hat{\lambda}_k}$ is the kth largest singular value of $\widehat{\mathbf{M}}$
        - $\hat{\lambda}_k$: the kth largest eigenvalue of $\widehat{\mathbf{M}}\widehat{\mathbf{M}}^\top$ (or $\widehat{\mathbf{M}}^\top\widehat{\mathbf{M}}$)
    - $\hat{\boldsymbol{a}}_k = \mathbf{S}_\mathbf{Y}^{-1/2} \hat{\boldsymbol{e}}_k$
        - $\hat{\boldsymbol{e}}_k$: the left-singular vector corresponding to the kth largest singular value of $\widehat{\mathbf{M}}$, i.e., the eigenvector corresponding to the kth largest eigenvalue of $\widehat{\mathbf{M}}\widehat{\mathbf{M}}^\top$
    - $\hat{\boldsymbol{b}}_k = \mathbf{S}_\mathbf{X}^{-1/2} \hat{\boldsymbol{f}}_k$
        - $\hat{\boldsymbol{f}}_k$: the right-singular vector corresponding to the kth largest singular value of $\widehat{\mathbf{M}}$, i.e., the eigenvector corresponding to the kth largest eigenvalue of $\widehat{\mathbf{M}}^\top\widehat{\mathbf{M}}$