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 $\mathbf{X}$
    * $p \le q$
    * $\mathbf{\Sigma_Y} = \mathrm{var}(\mathbf{Y}) > 0$, $\mathbf{\Sigma_X} = \mathrm{var}(\mathbf{X}) > 0$ 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: $(\boldsymbol{a}_k^\top \mathbf{Y}_C, \boldsymbol{b}_k^\top \mathbf{X}_C)$ with subscript C standing for centering
    * (The kth) canonical correlation: $\rho_k = \mathrm{corr}(\boldsymbol{a}_k^\top \mathbf{Y}, \boldsymbol{b}_k^\top \mathbf{X}) = \mathrm{corr}(\boldsymbol{a}_k^\top \mathbf{Y}_C, \boldsymbol{b}_k^\top \mathbf{X}_C)$
  - Goal: find $\boldsymbol{a}_k$ and $\boldsymbol{b}_k$, $k = 1, \ldots, r \le p$, 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 $\mathbf{M}^\top\mathbf{M}$

For $k=1$, to maximize $a_1^T \Sigma_{YX} b_1$ subject to $a_1^T \Sigma_Y a_1 = b_1^T \Sigma_X b_1 = 1$

i.e. maximize $\tilde{a}_1^T M \tilde{b}_1$ subject to $\tilde{a}_1^T \tilde{a}_1 = \tilde{b}_1^T \tilde{b}_1 = 1$,

where $\tilde{a}_1 = \Sigma_Y^{1/2} a_1$, $\tilde{b}_1 = \Sigma_X^{1/2} b_1$ & $M = \Sigma_Y^{-1/2} \Sigma_{YX} \Sigma_X^{-1/2}$

By the method of Lagrange multiplier,

take partial derivatives of $\tilde{a}_1^T M \tilde{b}_1 - \theta_1(\tilde{a}_1^T \tilde{a}_1 - 1) - \theta_2(\tilde{b}_1^T \tilde{b}_1 - 1)$,

then

$$\begin{cases} M\tilde{b}_1 - 2\theta_1 \tilde{a}_1 = 0 & ① \\ M^T \tilde{a}_1 - 2\theta_2 \tilde{b}_1 = 0 & ② \\ \tilde{a}_1^T \tilde{a}_1 = 1 \\ \tilde{b}_1^T \tilde{b}_1 = 1 \end{cases}$$

① $\Rightarrow M^T M \tilde{b}_1 - 2\theta_1 M^T \tilde{a}_1 = 0$

$\Rightarrow M^T M \tilde{b}_1 = 4\theta_1\theta_2 \tilde{b}_1$ (by ②)

$\Rightarrow \tilde{b}_1$ is the eigenvector corresponding to the 1st eigenvalue of $M^T M$
(refer to the proof for PCA in Lec 21)

Similarly, $\tilde{a}_1$ is the eigenvector corresponding to the 1st eigenvalue of MM^T

For $k=2$, to maximize $\tilde{a}_2^T M b_2$ subject to $\tilde{a}_2^T \tilde{a}_2 = \tilde{b}_2^T \tilde{b}_2 = 1$ & $\tilde{a}_2^T M \tilde{b}_1 = \tilde{a}_1^T M \tilde{b}_2 = 0$

where $\tilde{a}_2 = \Sigma_Y^{-1/2} a_2$ & $\tilde{b}_2 = \Sigma_X^{-1/2} b_2$

Refer again to the proof for PCA in Lec 21.

$\tilde{a}_2$ (resp. $\tilde{b}_2$) is the eigenvector corresponding to the 2nd eigenvalue of MM^T (resp. $M^T M$)

```r
install.packages("expm")
options(digits=4)
(Sigma_Y = matrix(c(1, 0.4, 0.4, 1), ncol = 2))
(Sigma_X = matrix(c(1, 0.2, 0.2, 1), ncol = 2))
(Sigma_YX = matrix(c(0.5, 0.3, 0.6, 0.4), ncol = 2))
(Sigma_XY = t(Sigma_YX))

Sigma_Y_sqrt = expm::sqrtm(Sigma_Y)
Sigma_X_sqrt = expm::sqrtm(Sigma_X)
M = solve(Sigma_Y_sqrt) %*% Sigma_YX %*% solve(Sigma_X_sqrt)

decomp1 = eigen(M %*% t(M))
a1 = solve(Sigma_Y_sqrt) %*% decomp1$vector[,1]
decomp2 = eigen(t(M) %*% M)
b1 = solve(Sigma_X_sqrt) %*% decomp2$vector[,1]
cbind(a1, b1) # the 1st pair of canonical directions
(rho1 = sqrt(decomp1$values[1])) # the 1st pair of canonical correlation
(rho1 = sqrt(decomp2$values[1])) # the 1st pair of canonical correlation

decomp3 = svd(M)
```

```r
a1 = solve(Sigma_Y_sqrt) %*% decomp3$u[,1]
b1 = solve(Sigma_X_sqrt) %*% decomp3$v[,1]
cbind(a1, b1) # the 1st pair of canonical directions
(rho1 = decomp3$d[1]) # the 1st pair of canonical correlation
```