High-dimensional data analysis

Lecture 4

Canonical Correlation Analysis

Innopolis University

Fall 2018

Principal Component Analysis

- Instead of looking at individual features, let’s look at their combinations
- How to choose these combinations?
  - Linear combinations,
  - that best explain the variability of data
- How many combinations to choose?
  - More combinations: higher accuracy
  - Less combinations: faster computations, better interpretability

Canonical Correlation Analysis

- We have two sets of features
  - Or, as an edge case, single set of features and a target variable
- We want to find linear combinations of features in each set that correlate strongly with each other.

Canonical Correlation Analysis

- Goals of Multivariate data analysis:
  - **understand the structure in the data;**
  - summarize data in simpler ways;
  - **find the relationship between parts of the data**;
  - make decisions based on the data.

Example

- Vehicle data from 1983 ASA Data Exposition
- Five variables:
  - Physical properties: Displacement, Horse power, Weight
  - Performance properties: Acceleration, Miles per gallon

- How are physical and performance properties related?

Theorem

- Let $A \in \mathbb{R}^{p \times q}$, $\text{rank } A = r$,

$$A = U\Lambda V^T; \qquad B = AA^T; \quad C = A^T A$$

- Then:
  - $\text{rank } B = \text{rank } C = r$
  - $B = UDU^T$; $C = VDV^T$, where $D = \Lambda^2$
  - Eigenvectors $\boldsymbol{u}_k$ of B and $\boldsymbol{v}_k$ of C have properties (for $k = 1 \ldots r$):
    - $A\boldsymbol{v}_k = \lambda_k \boldsymbol{u}_k$
    - $A^T \boldsymbol{u}_k = \lambda_k \boldsymbol{v}_k$

Normalization

- $$X \sim \left(\mu, \sigma^2\right) \rightarrow X_\Sigma = \frac{(X - \mu)}{\sigma} \sim (0, 1)$$

$$\boldsymbol{X} \sim (\boldsymbol{\mu}, \Sigma) \rightarrow \boldsymbol{X}_\Sigma = \Sigma^{-1/2}(\boldsymbol{X} - \boldsymbol{\mu}) \sim (\mathbf{0}, I)$$

Canonical Correlation Analysis

- $\boldsymbol{X}^{[1]} \sim (\boldsymbol{\mu}_1, \Sigma_1) \in \mathbb{R}^{d_1}, \boldsymbol{X}^{[2]} \sim (\boldsymbol{\mu}_2, \Sigma_2) \in \mathbb{R}^{d_2}$

- We want to find $\boldsymbol{a}_k \in \mathbb{R}^{d_1}, \boldsymbol{b}_k \in \mathbb{R}^{d_2}, k = 1..r$:
  - $\|\boldsymbol{a}_k\| = \|\boldsymbol{b}_k\| = 1, \ \left(\boldsymbol{a}_i, \boldsymbol{a}_j\right) = \left(\boldsymbol{b}_i, \boldsymbol{b}_j\right) = 0 \text{ for } i \neq j$
  - $U_k = \boldsymbol{a}_k^T \boldsymbol{X}_\Sigma^{[1]}, V_k = \boldsymbol{b}_k^T \boldsymbol{X}_\Sigma^{[2]}$
- Such that correlation between U_k and V_k decreases with k:
  - $|\mathrm{cov}(U_1, V_1)| \geq |\mathrm{cov}(U_2, V_2)| \geq \cdots \geq |\mathrm{cov}(U_r, V_r)| > 0$
  - $\mathrm{cov}(U_1, V_1)$ has maximum possible value

Canonical Correlation Matrix

- $\boldsymbol{X}^{[1]} \sim (\boldsymbol{\mu}_1, \Sigma_1) \in \mathbb{R}^{d_1}, \boldsymbol{X}^{[2]} \sim (\boldsymbol{\mu}_2, \Sigma_2) \in \mathbb{R}^{d_2}$
- $\Sigma_{12} = \operatorname{cov}\left(\boldsymbol{X}^{[1]}, \boldsymbol{X}^{[2]}\right)$

- $C = \operatorname{cov}\left(\boldsymbol{X}_\Sigma^{[1]}, \boldsymbol{X}_\Sigma^{[2]}\right) = \Sigma_1^{-1/2} \Sigma_{12} \Sigma_2^{-1/2}$ – canonical correlation matrix
- $R^{[C,1]} = CC^T, \quad R^{[C,2]} = C^T C$ – matrices of multivariate coefficients of determination

Theorem (the familiar one)

- Let $C \in \mathbb{R}^{p \times q}$, rank $C = r$,

$$C = U \Lambda V^T$$

- Then for $k = 1 \dots r$:
  - $C\boldsymbol{v}_k = \lambda_k \boldsymbol{u}_k$
  - $C^T \boldsymbol{u}_k = \lambda_k \boldsymbol{v}_k$

Canonical Correlation Analysis: 1st pair

- We want to maximize:

$$q = \mathrm{cov}(U_1, V_1) = \mathrm{cov}\left(\boldsymbol{a}_1^T \boldsymbol{X}_\Sigma^{[1]}, \boldsymbol{b}_1^T \boldsymbol{X}_\Sigma^{[2]}\right) = \boldsymbol{a}_1^T \mathrm{cov}\left(\boldsymbol{X}_\Sigma^{[1]}, \boldsymbol{X}_\Sigma^{[2]}\right) \boldsymbol{b}_1$$
$$= \boldsymbol{a}_1^T C \boldsymbol{b}_1 = (\boldsymbol{a}_1, C\boldsymbol{b}_1)$$

By Cauchy-Schwartz:

$$(\boldsymbol{a}_1, C\boldsymbol{b}_1)^2 \leq \left(\boldsymbol{a}_1 \boldsymbol{a}_1^T\right)\left(\boldsymbol{b}_1^T C^T C \boldsymbol{b}_1\right) = \boldsymbol{b}_1^T R^{[C,2]} \boldsymbol{b}_1$$

We get equality iff $\alpha \boldsymbol{a}_1 = C\boldsymbol{b}_1$ for some α

Canonical Correlation Analysis

- $\boldsymbol{X}^{[1]} \sim (\boldsymbol{\mu}_1, \Sigma_1) \in \mathbb{R}^{d_1}, \boldsymbol{X}^{[2]} \sim (\boldsymbol{\mu}_2, \Sigma_2) \in \mathbb{R}^{d_2}$
- $\Sigma_{12} = \mathrm{cov}\left(\boldsymbol{X}^{[1]}, \boldsymbol{X}^{[2]}\right), \quad C = \Sigma_1^{-1/2}\Sigma_{12}\Sigma_2^{-1/2} = P\Gamma Q^T$

- $U_k = \boldsymbol{p}_k^T \boldsymbol{X}_\Sigma^{[1]}, V_k = \boldsymbol{q}_k^T \boldsymbol{X}_\Sigma^{[2]}$ – canonical correlation (CC) scores
- $\boldsymbol{U}^{(k)} = [U_1 \ \ U_2 \ \ \ldots U_k]^T, \boldsymbol{V}^{(k)} = [V_1 \ \ V_2 \ \ \ldots V_k]^T$ – vectors of CCs
- $\boldsymbol{\varphi}_k = \Sigma_1^{-1/2}\boldsymbol{p}_k, \boldsymbol{\psi}_k = \Sigma_2^{-1/2}\boldsymbol{q}_k$ – CC transforms

Sample Canonical Correlation

- $\mathbb{X}^{[1]} = \left[\boldsymbol{X}_1^{[1]}, \dots, \boldsymbol{X}_n^{[1]}\right] \in \mathbb{R}^{d_1 \times n}, \mathbb{X}^{[2]} = \left[\boldsymbol{X}_1^{[2]}, \dots, \boldsymbol{X}_n^{[2]}\right] \in \mathbb{R}^{d_2 \times n}$
- $\mathbb{X} = \begin{bmatrix} \mathbb{X}^{[1]} \\ \mathbb{X}^{[2]} \end{bmatrix} \sim \mathrm{Sam}(\overline{\boldsymbol{X}}, S) = \mathrm{Sam}\left(\begin{bmatrix} \overline{\boldsymbol{X}_1} \\ \overline{\boldsymbol{X}_2} \end{bmatrix}, \begin{pmatrix} S_{11} & S_{12} \\ S_{12}^T & S_{22} \end{pmatrix}\right)$
  - $s_{ij} = \frac{1}{n-1} \sum_{k=1}^{n} \left(\boldsymbol{X}_k^{[i]} - \overline{\boldsymbol{X}_i}\right)\left(\boldsymbol{X}_k^{[j]} - \overline{\boldsymbol{X}_j}\right)^T = \frac{1}{n-1} \mathbb{X}_0^{[i]} \left(\mathbb{X}_0^{[j]}\right)^T$
- $C = S_{11}^{-1/2}\, S_{12}\, S_{22}^{-1/2}$ – sample canonical correlation matrix
  - $C = P\Gamma Q^T$
- $\boldsymbol{U}_{\blacksquare k} = \boldsymbol{p}_k^T \mathbb{X}_{\mathrm{S}}^{[1]}, \boldsymbol{V}_{\blacksquare k} = \boldsymbol{q}_k^T \mathbb{X}_{\mathrm{S}}^{[2]}$ – pair of canonical correlation vectors

Properties

- $\boldsymbol{X}^{[1]} \sim (\boldsymbol{\mu}_1, \Sigma_1) \in \mathbb{R}^{d_1}, \boldsymbol{X}^{[2]} \sim (\boldsymbol{\mu}_2, \Sigma_2) \in \mathbb{R}^{d_2}$
- $\Sigma_{12} = \mathrm{cov}\left(\boldsymbol{X}^{[1]}, \boldsymbol{X}^{[2]}\right), \quad C = \Sigma_1^{-1/2}\Sigma_{12}\Sigma_2^{-1/2} = P\Gamma Q^T$
- U_k, V_l – CC scores, $\boldsymbol{U}^{(k)}, \boldsymbol{V}^{(l)}$ – CC vectors

- $E\begin{bmatrix}\boldsymbol{U}^{(k)} \\ \boldsymbol{V}^{(l)}\end{bmatrix} = \begin{bmatrix}\mathbf{0} \\ \mathbf{0}\end{bmatrix}, \;\; \mathrm{Var}\begin{bmatrix}\boldsymbol{U}^{(k)} \\ \boldsymbol{V}^{(l)}\end{bmatrix} = \begin{bmatrix} I & \Gamma_{k\times l} \\ \Gamma_{k\times l}^T & I\end{bmatrix}$
- $\mathrm{Var}(U_k) = \mathrm{Var}(V_l) = 1$
- $\mathrm{Cov}(U_k, V_k) = \pm\,\gamma_k, \mathrm{Cov}(U_k, V_l) = 0 \; (k \neq l)$

CCs and Linear Transformations

- $\boldsymbol{X}^{[1]} \sim (\boldsymbol{\mu}_1, \Sigma_1) \in \mathbb{R}^{d_1}, \boldsymbol{X}^{[2]} \sim (\boldsymbol{\mu}_2, \Sigma_2) \in \mathbb{R}^{d_2}$
- $\boldsymbol{T}^{[1]} = A_1 \boldsymbol{X}^{[1]} + \boldsymbol{a}_1, \boldsymbol{T}^{[2]} = A_2 \boldsymbol{X}^{[2]} + \boldsymbol{a}_2$
- $C_X = C\left(\boldsymbol{X}^{[1]}, \boldsymbol{X}^{[2]}\right) = P_X \Gamma_X Q_X, C_T = C\left(\boldsymbol{T}^{[1]}, \boldsymbol{T}^{[2]}\right) = P_T \Gamma_T Q_T$

- $\Gamma_T = \Gamma_X$
- $\boldsymbol{p}_{T,k} = \left(A_1 \Sigma_1 A_1^T\right)^{1/2} \left(A_1^T\right)^{-1} \Sigma_1^{-1/2} \boldsymbol{p}_{X,k}$
- $\boldsymbol{q}_{T,k} = \left(A_2 \Sigma_2 A_2^T\right)^{1/2} \left(A_2^T\right)^{-1} \Sigma_2^{-1/2} \boldsymbol{q}_{X,k}$
- $U_{T,k} = \boldsymbol{p}_k^T \Sigma_1^{-1/2} \left(\boldsymbol{X}^{[1]} - \boldsymbol{\mu}_1\right), U_{T,k} = \boldsymbol{q}_k^T \Sigma_2^{-1/2} \left(\boldsymbol{X}^{[2]} - \boldsymbol{\mu}_2\right)$