Sparsity preserved canonical correlation analysis

Presenter: Muhammad Ali Qadar
Authors: Abd-Karim Seghouane and Muhammad Ali Qadar

Regular Session, Paper # 3200
Department of Electrical and Electronic Engineering,
The University of Melbourne

THE UNIVERSITY OF MELBOURNE

Outline

Introduction

- Canonical correlation analysis (CCA) describes relationship between two set of variables.
- Sparse CCA improves interpretability in high-dimensional settings.
- However, sparse CCA do not preserve sparsity across canonical directions[1].
- We propose a sparse CCA that ensures consistency in recovered sparsity patterns.
- A penalized alternating least squares framework with l_2^1-norm to enforce block sparsity.

---

[1] A. Seghouane, N. Shokouhi and I. Koch. "Sparse Principal Component Analysis With Preserved Sparsity Pattern". In: *IEEE Transactions on Image Processing* 28.7 (2019), pp. 3274–3285. ISSN: 1057-7149. DOI: 10.1109/TIP.2019.2895464.

Background

Canonical correlation analysis

Consider N i.i.d samples of $\mathbf{x} \in \mathbb{R}^p$ and $\mathbf{y} \in \mathbb{R}^q$, CCA extracts the pairs of vectors $\mathbf{w}_{x_i} \in \mathbb{R}^p$ and $\mathbf{w}_{y_i} \in \mathbb{R}^q$, $1 \leq i \leq d \leq \min(p, q)$ such that

$$
\begin{aligned}
(\mathbf{w}_{x_i}, \mathbf{w}_{y_i}) &= \underset{\mathbf{w}_x, \mathbf{w}_y}{\arg\max}\, \mathbf{w}_x{}^\top \mathbf{\Sigma}_{xy} \mathbf{w}_y \\
\text{subject to} \quad & \mathbf{w}_x{}^\top \mathbf{\Sigma}_x \mathbf{w}_x = 1, \quad \mathbf{w}_y{}^\top \mathbf{\Sigma}_y \mathbf{w}_y = 1 \\
& \mathbf{w}_x^\top \mathbf{\Sigma}_x \mathbf{w}_{x_j} = 0, \quad \mathbf{w}_y^\top \mathbf{\Sigma}_y \mathbf{w}_{y_j} = 0 \\
& \forall\, 1 \leq j \leq d-1,\ j < i
\end{aligned} \tag{1}
$$

where $\mathbf{\Sigma}_x = \mathbb{E}[\mathbf{x}^\top \mathbf{x}]$, $\mathbf{\Sigma}_y = \mathbb{E}[\mathbf{y}^\top \mathbf{y}]$, $\mathbf{\Sigma}_{xy} = \mathbf{\Sigma}_{yx}^\top = \mathbb{E}[\mathbf{x}^\top \mathbf{y}]$ are covariance and cross-covariances, respectively.

Background

Sparse CCA

When $p >> N$ and $q >> N$, it is likely that only a fraction of variables are informative. To address this selection problem, additional constraints has been imposed on CCA problem (1):

$$(\mathbf{w}_x, \mathbf{w}_y) = \arg\max_{\mathbf{w}_x, \mathbf{w}_y} \mathbf{w}_x^\top \mathbf{C}_{xy} \mathbf{w}_y \quad \text{subject to} \quad \|\mathbf{w}_x\|_2 \leq 1, \quad \|\mathbf{w}_y\|_2 \leq 1, \quad \|\mathbf{w}_x\|_1 \leq c_1, \quad \|\mathbf{w}_y\|_1 \leq c_2 \tag{2}$$

where c_1 and c_2 are regularization parameters.

Proposed method

The proposed method preserves sparsity among d sets of $\mathbf{W}_x \in \mathbb{R}^{p \times d}$ and $\mathbf{W}_y \in \mathbb{R}^{q \times d}$ by imposing l_2^1-norm group sparse penality. The proposed objective function is:

$$\min_{\mathbf{W}_x, \mathbf{W}_y} \|\mathbf{X}\mathbf{W}_x - \mathbf{Y}\mathbf{W}_y\|_F^2 + \alpha \sum_{i=1}^{p} \left\|\mathbf{w}_x^i\right\|_2 + \beta \sum_{j=1}^{q} \left\|\mathbf{w}_y^j\right\|_2 \quad (3)$$

where $\mathbf{w}_x^i$ and $\mathbf{w}_y^j$ are both $d \times 1$ vectors corresponding to the i^{th} and j^{th} rows of $\mathbf{W}_x$ and $\mathbf{W}_y$, respectively. α and β are sparsity controlling parameters.

Proposed method

$\mathbf{W}_x$ and $\mathbf{W}_y$ are obtained by a block coordinate descent method where each of the variables are computed row by row using the closed form solutions.

$$\mathbf{w}_y^j = \frac{1}{\mathbf{y}_j^\top \mathbf{y}_j} \left[1 - \frac{\beta}{2 \left\| \mathbf{y}_j^\top \mathbf{E}_j \right\|_2} \right]_+ \mathbf{y}_j^\top \mathbf{E}_j \quad (4)$$

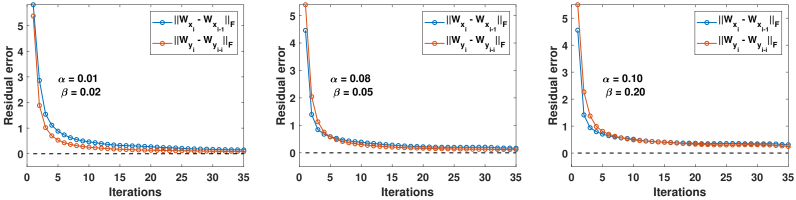
where $\mathbf{y}_j$ is the j^{th} column of $\mathbf{Y}$, $\mathbf{E}_j = \mathbf{X}\mathbf{W}_x - \sum_{\substack{i=1 \\ i \neq j}}^{q} \mathbf{y}_i \mathbf{w}_y^i$,

$[x]_+ = max(0, x)$, and

$$\mathbf{w}_x^i = \frac{1}{\mathbf{x}_i^\top \mathbf{x}_i} \left[1 - \frac{\alpha}{2 \left\| \mathbf{x}_i^\top \mathbf{F}_i \right\|_2} \right]_+ \mathbf{x}_i^\top \mathbf{F}_i \quad (5)$$

where $\mathbf{x}_i$ is the i^{th} column of $\mathbf{X}$ and $\mathbf{F}_i = \sum_{\substack{j=1 \\ j \neq i}}^{p} \mathbf{x}_j \mathbf{w}_x^j - \mathbf{Y}\mathbf{W}_y$.

Proposed method



(a) (b) (c)

Figure: Shows changes in estimated canonical loading vectors for the proposed algorithm in terms of Frobenius norm.

Experimental results

Simulation

To generate data matrices $(\mathbf{X}, \mathbf{Y})$

- Canonical projection matrices $\mathbf{W}_x \in \mathbb{R}^{p_x \times q_x}$ and $\mathbf{W}_y \in \mathbb{R}^{q_x \times q_y}$
- $\mathbf{D}_x = diag(vec(\mathbf{W}_x))$ and $\mathbf{D}_y = diag(vec(\mathbf{W}_y))$, $\mathbf{D}_x \in \mathbb{R}^{p_x q_x \times p_x q_x}$, $\mathbf{D}_y \in \mathbb{R}^{p_y q_y \times p_y q_y}$
- $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}), \mathbf{z} \in \mathbb{R}^d$ with $(d << p, d << q)$
- Orthonormal basis matrix $\mathbf{U}_x$ of size $p_x q_x \times d$

Experimental results

Simulation

- $\mathbf{x} = \mathbf{D}_x\mathbf{U}_x\mathbf{z} + \boldsymbol{\eta}_x$, $\boldsymbol{\eta}_x \sim (\mathbf{0}, \sigma_x^2\mathbf{I})$ with $\sigma_x = 0.15$ (and $\sigma_y = 0.1$), $\mathbf{x} \in \mathbb{R}^{p_x q_x}$
- Generate N samples $\mathbf{X} = [\mathbf{x}_1^\top \cdots \mathbf{x}_N^\top]$ and $\mathbf{Y} = [\mathbf{y}_1^\top \cdots \mathbf{y}_N^\top]$
- We set $N = 150$, $p_x = q_x = 15$, and $p_y = q_y = 14$.
- The value of $d = 5$ set using the model selection criterion[a]

[a] A. K. Seghouane and N. Shokouhi. "Estimating the Number of Significant Canonical Coordinates". In: *IEEE Access* 7 (2019), pp. 108806–108817. ISSN: 2169-3536.

Experimental results

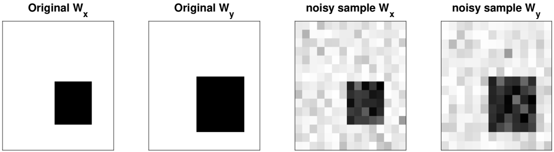


Figure: Completely overlapping sparsity pattern of the original canonical loading vectors.

Experimental results

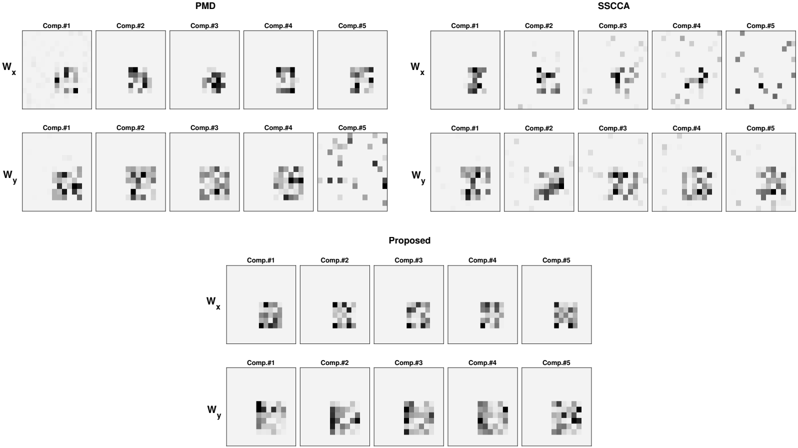


Figure: Shows consistency in sparsity across estimated $\mathbf{W}_x$ and $\mathbf{W}_y$.

Experimental results

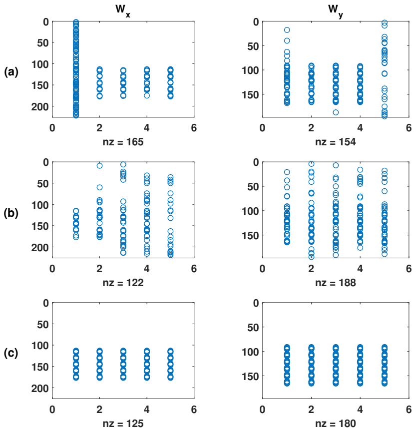


Figure: Sparse grid for the recovered sparsity patterns across $\mathbf{W}_x$ and $\mathbf{W}_y$.

Conclusion

- An SCCA algorithm that share same sparsity across components.
- l_2^1-norm penalization was used to encourage block sparsity.
- Experiments show strength of the proposed algorithm in terms of estimating consistent sparsity patterns.

References

1. A. Seghouane, N. Shokouhi and I. Koch. "Sparse Principal Component Analysis With Preserved Sparsity Pattern". In: *IEEE Transactions on Image Processing* 28.7 (2019), pp. 3274–3285. ISSN: 1057-7149. DOI: 10.1109/TIP.2019.2895464
2. A. K. Seghouane and N. Shokouhi. "Estimating the Number of Significant Canonical Coordinates". In: *IEEE Access* 7 (2019), pp. 108806–108817. ISSN: 2169-3536
3. D. M. Witten, R. Tibshirani and T. Hastie. "A penalized matrix decomposition, with applications to sparse principal components and canonical correlation analysis". In: *Biostatistics* 10.3 (2009), pp. 515–534
4. Xi Chen, Liu Han and Jaime Carbonell. "Structured Sparse Canonical Correlation Analysis". In: *Proceedings of the Fifteenth International Conference on Artificial Intelligence and Statistics* (2012), pp. 199–207

Thank You!