A Principal Component Analysis for Ordinal Data

Hugo Alonso [1,2][0000-0002-1599-5392] and Adelaide Freitas [2,3][0000-0002-4685-1615]
hugo.alonso@upt.pt, adelaide@ua.pt

[1] *Portucalense University/Research on Economics, Management and Information Technologies (REMIT), Rua Dr. António Bernardino de Almeida, 541, 4200-072 Porto, Portugal*

[2] *Center for Research and Development in Mathematics and Applications (CIDMA), Department of Mathematics, University of Aveiro, 3810-193 Aveiro, Portugal*

[3] *Department of Mathematics, University of Aveiro, Campus de Santiago, 3810-193 Aveiro, Portugal*

Abstract: Large datasets are common and can be challenging to interpret. Principal Component Analysis (PCA) is a technique used to reduce the dimensionality of such datasets, making them easier to understand while preserving as much of the original information as possible [1]. In some cases, datasets consist solely of ordinal variables measured on the same scale. A common example is job interviews, where candidates are evaluated across multiple criteria using the same Likert scale. In this work, we propose a PCA approach specifically designed for ordinal data measured on the same scale. The idea is to replace Pearson's correlation matrix, used in standard PCA for quantitative variables, with a matrix based on r_{int}, a measure of association between ordinal variables introduced in [2, 3]. Although r_{int} can be computed regardless of whether the ordinal variables share the same scale or not, the calculation formula is simpler when the scales are equal, making it more practical to use in such cases. We present results from an empirical study in which we compare the performance of our r_{int}-based PCA with PCA based on Spearman's and Kendall's correlation coefficients, using several datasets consisting exclusively of ordinal variables measured on the same scale.

Keywords: Ordinal data · Principal Component Analysis · Rank correlation

Acknowledgements: This work is partially supported by CIDMA under the Portuguese Foundation for Science and Technology (FCT, https://ror.org/00snfqn58) Multi-Annual Financing Program for R&D Units.

References

[1] Jolliffe, I.T., Cadima, J.: Principal component analysis: a review and recent developments. Philosophical Transactions of the Royal Society A, **374**, pages 1-16 (2016). `DOI:10.1098/rsta.2015.0202`

[2] Pinto da Costa, J.F., Alonso, H., Cardoso, J.S.: The unimodal model for the classification of ordinal data. Neural Networks, **21**, pages 78-91 (2008). `DOI:110.1016/j.neunet.2007.10.003`

[3] Pinto da Costa, J.F., Alonso, H., Cardoso, J.S.: Corrigendum to "The unimodal model for the classification of ordinal data" [Neural Netw. 21 (2008) 78-79]. Neural Networks, **59**, pages 73-75 (2014). `DOI:10.1016/j.neunet.2014.06.003`